

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 23, 2001, 14:20:36 ; Search time 11.93 Seconds

(without alignments)
1263.404 Million cell updates/sec

Title: US-08-883-036a-2

Perfect score: 440

Sequence: 1 MEGRGQNPASGARKRRHP.....HLSSGKFMYLEGNADSAMS 440

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size: 6

Total number of hits satisfying chosen parameters: 843

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.8	598	1 NU5M_BRAFLA	079422 branchiosto
2	8	1.8	599	1 NU5M_BRAFLA	047430 branchiosto
3	8	1.8	1239	1 V120_EBV	P03189 epstein-bar
4	7	1.6	61	1 AKH2_LOGMI	P08379 locusta midg
5	7	1.6	79	1 RL27_PUPAP	044161 prunus arme
6	7	1.6	111	1 SINR_BACSU	P06533 bacillus su
7	7	1.6	169	1 GP38_CANFA	Q95152 canis fami
8	7	1.6	174	1 NUC_SHIFL	P29769 shigella fl
9	7	1.6	179	1 RIMM_MYCLE	033016 mycobacteri
10	7	1.6	197	1 YDEQ_BACSU	P34677 caenorhabd
11	7	1.6	243	1 Y02T_CAELU	000058 utromyces fa
12	7	1.6	256	1 MTDA_UROFA	P37306 escherichia
13	7	1.6	297	1 ARCC_ECOLI	026869 methanobact
14	7	1.6	308	1 METC_METTH	P52024 streptomyce
15	7	1.6	319	1 MERR_STRLI	P41035 oryctolagus
16	7	1.6	328	1 HOLB_PSEAE	043570 homo sapien
17	7	1.6	333	1 IF2B_RABIT	075683 homo sapien
18	7	1.6	354	1 CAHC_HUMAN	068640 corynebacte
19	7	1.6	361	1 SURE_HUMAN	P45100 rhizobium m
20	7	1.6	379	1 METX_CORGL	099XZ7 mus musculu
21	7	1.6	381	1 CYCH_RHIME	P27782 mus musculu
22	7	1.6	395	1 NR23_MOUSE	Q99XN1 rattus norv
23	7	1.6	397	1 LEPI_MOUSE	Q9UJ12 homo sapien
24	7	1.6	397	1 LEPI_MOUSE	004556 arabidopsi
25	7	1.6	399	1 LEPI_MOUSE	026328 methanobact
26	7	1.6	435	1 LEPI_MOUSE	092180 helicobacte
27	7	1.6	437	1 SYD_METTH	P51906 mus musculu
28	7	1.6	437	1 RECN_HELPJ	P51907 rattus norv
29	7	1.6	523	1 EAT3_MOUSE	P43005 homo sapien
30	7	1.6	523	1 EAT3_MOUSE	059573 mycobacteri
31	7	1.6	524	1 EAT3_MOUSE	P80461 nicotiana t
32	7	1.6	539	1 CHG1_MYCTU	
33	7	1.6	557	1 GSHC_TOBAC	

107	6	1.4	152	1	RL9_MYCLE	P46385 mycobacteri	180	6	1.4	237	1	CD63_RABIT	Q28709 oryctolagus
108	6	1.4	154	1	HS11_LYCES	P30221 lycopersico	181	6	1.4	238	1	YS76_CAEEL	Q09619 caenorhabdi
109	6	1.4	154	1	HS15_SOYBN	P04995 glycerine max	182	6	1.4	237	1	YEEN_ECOLI	P76351 escherichia
110	6	1.4	156	1	HS11_ARATH	P19936 arabidopsis	183	6	1.4	241	1	FADR_HABIN	P44705 nemophilus
111	6	1.4	158	1	HS12_MEDSA	P27880 medicago sa	184	6	1.4	244	1	TRUA_MYCCE	P47428 mycoplasma
112	6	1.4	160	1	TATB_HELPJ	Q92m58 helicobacte	185	6	1.4	245	1	YIPA_YEREN	P27461 yersinia en
113	6	1.4	160	1	YGAD_PSEPU	P72227 pseudomonas	186	6	1.4	246	1	CFL_ARATH	P41088 arabidopsis
114	6	1.4	161	1	HS16_SOYBN	P05478 glycerine max	187	6	1.4	247	1	CAH_METTE	P40881 methanosarc
115	6	1.4	162	1	HXA9_CAVPO	P51783 cavia porce	188	6	1.4	247	1	EFTS_SPIPL	P34828 spirulina p
116	6	1.4	162	1	Y096_GVCL	P41728 cryptophab	189	6	1.4	247	1	PGG2_MASLA	P29372 mastigoclad
117	6	1.4	163	1	YAG5_SCHPO	Q09669 schizosacch	190	6	1.4	249	1	COBS_MYCTU	Q10397 mycobacteri
118	6	1.4	164	1	GP38_ERWCA	P31709 erwina car	191	6	1.4	252	1	PMN_CANAL	P31353 candida alb
119	6	1.4	166	1	GSPM_ERWCA	Q64994 rattus norv	192	6	1.4	252	1	PSA6_YEAST	P21243 saccharomyc
120	6	1.4	166	1	Y410_SYNY3	Q05732 synechocyst	193	6	1.4	253	1	UDE_SALTY	Q33808 salmonella
121	6	1.4	168	1	SSH_BRUAB	Q07436 bruceella ab	194	6	1.4	254	1	UDE_KREAE	Q08444 klebsiella
122	6	1.4	169	1	AROK_CORGL	Q9x5d1 corynebacte	195	6	1.4	254	1	FC3A_HUMAN	P08637 homo sapien
123	6	1.4	172	1	ESP4_LACVY	P35757 laetitia viv	196	6	1.4	255	1	VSPB_SOYBN	P10743 glycine max
124	6	1.4	172	1	Y074_NPVOP	O10327 orydia pseu	197	6	1.4	255	1	E128_DROME	P08761 dirosophila
125	6	1.4	177	1	IMP2_YEAST	P46972 saccharomyc	198	6	1.4	255	1	ETEB_HUMAN	P38117 homo sapien
126	6	1.4	179	1	ATPF_SYNY3	P27181 synechocyst	199	6	1.4	255	1	HCD2_DROME	O18404 dirosophila
127	6	1.4	179	1	VG59_HSVBR	P28983 equine herp	200	6	1.4	255	1	HVCG_ECOLI	P16433 escherichia
128	6	1.4	180	1	PORC_METTH	Q27772 methanobact	201	6	1.4	255	1	TATC_AZOCH	P54085 azotobacter
129	6	1.4	180	1	SMCX_CRIGR	P41728 cricetulus	202	6	1.4	256	1	YDHL_HSVSC	P22575 herpesvirus
130	6	1.4	181	1	NDKM_COLLI	P87355 columba liv	203	6	1.4	257	1	ECBH_MYCLE	O07137 mycobacteri
131	6	1.4	185	1	GRA2_TOXGO	O58411 p pyruvate/	204	6	1.4	257	1	ECBH_MYCTU	O53418 mycobacteri
132	6	1.4	185	1	PORC_PYRHO	O86770 streptomyc	205	6	1.4	257	1	ECBH_RHIME	Q52995 rhizobium m
133	6	1.4	185	1	RRF_STRICO	Q9xth9 arabidopsis	206	6	1.4	257	1	OMPV_VIBCH	P06111 vibrio chol
134	6	1.4	185	1	THF2_ARATH	P02398 spinacia ol	207	6	1.4	259	1	PMW_SCHPO	Q9ut12 schizosacch
135	6	1.4	189	1	RK12_SPIOL	P30361 bradyrhizob	208	6	1.4	259	1	TRUA_METTA	O59069 methanococc
136	6	1.4	190	1	ISP2_BRAJA	P56940 rhodobacter	209	6	1.4	262	1	S3AD_KLEBN	P08881 klebsiella
137	6	1.4	191	1	COX2_RHOSH	O93118 anopheles g	210	6	1.4	263	1	S3AD_ECOLI	P04826 escherichia
138	6	1.4	192	1	AL6_ANOGA	P02606 gallus gall	211	6	1.4	264	1	COLI_MACNE	P01201 m corticotr
139	6	1.4	193	1	MLEBC_CHICK	P32293 phaseolus a	212	6	1.4	264	1	TRPC_LACLA	Q01999 lactococcus
140	6	1.4	194	1	AXA2_PHRAN	O27793 methanobact	213	6	1.4	264	1	YXEM_BACSU	P54952 bacillus su
141	6	1.4	194	1	KTHT_METTH	P08590 homo sapien	214	6	1.4	266	1	KLR8_MOUSE	O60682 mus musculu
142	6	1.4	194	1	MLEV_HUMAN	P09867 bos taurus	215	6	1.4	267	1	FTSQ_RICPR	Q92d85 rickettsia
143	6	1.4	195	1	ROAL_BOVIN	P37006 serratia ma	216	6	1.4	269	1	RAD_HUMAN	P53043 rattus norv
144	6	1.4	196	1	LAST_SERMA	P52815 homo sapien	217	6	1.4	269	1	PRYM_BACSP	Q99405 bacillus sp
145	6	1.4	198	1	RM12_HUMAN	P16409 rattus norv	218	6	1.4	269	1	SUBB_BACLE	P29599 bacillus le
146	6	1.4	199	1	TOEB_RAT	O28249 archaeoglob	219	6	1.4	269	1	SUBS_BACLE	P29600 bacillus le
147	6	1.4	201	1	NDCL_RHIME	O68854 rhizobium m	220	6	1.4	269	1	TRCL_STRCO	O68814 streptomyc
148	6	1.4	201	1	FLAI_ARCFU	O29208 archaeoglob	221	6	1.4	270	1	MER1_YEAST	P16523 saccharomyc
149	6	1.4	203	1	RM12_CRICR	P52827 cricetus cr	222	6	1.4	271	1	ECEL_CANAL	Q07730 candida alb
150	6	1.4	203	1	GIDB_COXBU	P94614 coxiella bu	223	6	1.4	272	1	E2F6_MOUSE	O54917 mus musculu
151	6	1.4	204	1	EMBP_CORIG	P81448 cricetulus	224	6	1.4	273	1	ROCL_NICGX	Q08935 nicotiana s
152	6	1.4	206	1	EMBP_CRIGR	P24407 homo sapien	225	6	1.4	274	1	Y975_TREPA	O83940 treponema p
153	6	1.4	207	1	RAB8_HUMAN	O69162 bradyrhizob	226	6	1.4	276	1	KLKA_HUMAN	O43240 homo sapien
154	6	1.4	208	1	ENGB_BRAJA	P24253 escherichia	227	6	1.4	276	1	SSUC_BACSU	P44011 bacillus su
155	6	1.4	210	1	YIHA_ECOLI	P24253 escherichia	228	6	1.4	277	1	CYST_ECOLI	P16701 escherichia
156	6	1.4	211	1	PDOL_ECOLI	P32174 escherichia	229	6	1.4	279	1	DAVE_SYNY3	P74667 synechocyst
157	6	1.4	211	1	RSJA_ARCFU	O27664 archaeoglob	230	6	1.4	279	1	ROCL_NICPL	P49313 nicotiana p
158	6	1.4	212	1	HIS5_NEIMA	O914h3 neisseria m	231	6	1.4	279	1	UDP_KLEPN	P52671 klebsiella
159	6	1.4	212	1	UL45_HSVTH	P18536 turkey herp	232	6	1.4	279	1	Y144_MYCCE	P47390 mycoplasma
160	6	1.4	212	1	CPPB_NEIGO	P07049 neisseria g	233	6	1.4	283	1	CXBI_HUMAN	P06034 homo sapien
161	6	1.4	213	1	EFIB_CANAL	P78590 candida alb	234	6	1.4	283	1	CXBI_MOUSE	P28230 mus musculu
162	6	1.4	213	1	NCCN_ALCXX	O44587 alcaligenes	235	6	1.4	284	1	CXBI_BOVIN	O18968 bos taurus
163	6	1.4	213	1	MGML_HUMAN	O9x5n5 homo sapien	236	6	1.4	284	1	STX2_CAEEL	Q20574 caenorhabdi
164	6	1.4	215	1	YD05_SCHPO	O14198 schizosacch	237	6	1.4	285	1	PYRI_MASLA	P11398 mastigoclad
165	6	1.4	215	1	CYTA_CLOAB	P33752 clostridium	238	6	1.4	285	1	PYRI_ANASP	P07123 anabaena sp
166	6	1.4	218	1	YBGJ_ECOLI	P75744 escherichia	239	6	1.4	286	1	ALF2_BACSU	P42420 bacillus su
167	6	1.4	218	1	YIAD_ECOLI	Q50664 mycobacteri	240	6	1.4	286	1	ATPF_BACSU	P57123 buchnera ap
168	6	1.4	219	1	EMBP_MYCTU	P13727 homo sapien	241	6	1.4	290	1	VP38_HSWAG	P306023 marek's dis
169	6	1.4	222	1	EMBP_HUMAN	Q9ubt3 homo sapien	242	6	1.4	290	1	YPIA_BACSU	Q09504 caenorhabdi
170	6	1.4	224	1	CDSN_PTIG	O19084 sus scrofa	243	6	1.4	290	1	Y012_CAEEL	P24037 pseudomonas
171	6	1.4	225	1	YB67_AERPE	Q9yuc5 aeropyrum p	244	6	1.4	291	1	C552_PSEST	O05190 culdabacter
172	6	1.4	225	1	COX2_ADABI	P43163 streptomyc	245	6	1.4	293	1	PARB_CANDR	Q00940 plicha angu
173	6	1.4	229	1	NMDA_STRPO	O03733 hypnale hyp	246	6	1.4	295	1	YHYA_PSESN	O01265 pseudomonas
174	6	1.4	233	1	FC3B_HUMAN	P42239 bacillus su	247	6	1.4	297	1	BMRU_BACSU	P39074 bacillus su
175	6	1.4	233	1	YCBG_BACSU	P08962 homo sapien	248	6	1.4	297	1	STX4_HUMAN	Q12846 homo sapien
176	6	1.4	237	1			249	6	1.4				
177	6	1.4					250	6	1.4				
178	6	1.4					251	6	1.4				
179	6	1.4					252	6	1.4				

253	6	1.4	297	1	YN24_CAEEL	P45964	caenorhabdi.	326	6	1.4	359	1	GCPE_HELPY	O25342	helicobacte
254	6	1.4	299	1	SC5D_HUMAN	O75845	homo sapien	327	6	1.4	361	1	CHSA_IPOCO	P48393	ipomoea cor
255	6	1.4	301	1	FMDC_METBA	O48943	methanosarc	328	6	1.4	361	1	DBPA_RAT	O62764	rattus norv
256	6	1.4	302	1	ATPG_STRLI	P50007	streptomyce	329	6	1.4	361	1	PDA6_ARATH	O22263	arabidopsis
257	6	1.4	302	1	PR12_DROME	P12982	drosophila	330	6	1.4	361	1	YUQO_ECOLI	P39341	escherichia
258	6	1.4	302	1	PR13_DROME	O05547	drosophila	331	6	1.4	362	1	CHSA_IPOPL	P48400	ipomoea pia
259	6	1.4	304	1	GAT1_CHICK	P17678	gallus gall	332	6	1.4	362	1	CHSA_IPOPL	P48401	ipomoea trl
260	6	1.4	304	1	Y245_SYNY3	P72702	synecocyst	333	6	1.4	362	1	KLFI_HUMAN	O13351	homo sapien
261	6	1.4	307	1	BRAD_PSEAE	P21627	pseudomonas	334	6	1.4	363	1	AROC_PSEAE	O91344	pseudomonas
262	6	1.4	307	1	VP53_BPAPS	O91125	bacteriophag	335	6	1.4	363	1	YGCC_ECOLI	O46899	escherichia
263	6	1.4	308	1	Y880_METJA	O58280	methanococc	336	6	1.4	365	1	CMLE_NEUCR	P38677	neurospora
264	6	1.4	309	1	MAT1_MOUSE	P51949	mus muscucu	337	6	1.4	367	1	CCR3_MOUSE	O88410	mus muscucu
265	6	1.4	314	1	PIX1_HUMAN	P78317	homo sapien	338	6	1.4	367	1	Y797_METJA	O58287	methanococc
266	6	1.4	315	1	O3A3_HUMAN	P47888	homo sapien	339	6	1.4	368	1	CCR3_HUMAN	P49682	homo sapien
267	6	1.4	315	1	PIX1_MOUSE	P70314	mus muscucu	340	6	1.4	368	1	ZN24_HUMAN	P17928	homo sapien
268	6	1.4	317	1	MSHR_ALCAA	P56442	alces alces	341	6	1.4	371	1	ROAL_HUMAN	P09551	homo sapien
269	6	1.4	317	1	MSHR_BOVIN	P47798	bos taurus	342	6	1.4	373	1	LEU3_CANAL	P87186	candida alb
270	6	1.4	317	1	MSHR_CANFA	O77616	canis famli	343	6	1.4	374	1	FLA1_BARBA	P35633	bartonella
271	6	1.4	317	1	MSHR_CAPCA	P56443	capreolus c	344	6	1.4	375	1	KMOS_MSVMH	P07331	moloney mur
272	6	1.4	317	1	MSHR_CAPII	P56444	capra hircu	345	6	1.4	375	1	NYAR_MOUSE	O61041	mus muscucu
273	6	1.4	317	1	MSHR_CEREL	P56445	cervus elap	346	6	1.4	376	1	GBAS_LYMSI	P30684	lymnaea sta
274	6	1.4	317	1	MSHR_DAMDA	P56446	dama dama	347	6	1.4	376	1	KMOS_MSVMH	P00537	moloney mur
275	6	1.4	317	1	MSHR_OVIMO	P56447	ovidos mosc	348	6	1.4	376	1	PT16_HUMAN	P35237	homo sapien
276	6	1.4	317	1	MSHR_RANTA	P56448	rangifer ta	349	6	1.4	377	1	HSE7_ARATH	O910d3	arabidopsis
277	6	1.4	317	1	MSHR_SHEEP	O19037	ovis aries	350	6	1.4	377	1	RECA_STRVL	P48295	streptomyce
278	6	1.4	317	1	MSHR_VULVU	O29154	vulpes vulp	351	6	1.4	378	1	ELYA_BACSP	P20724	bacillus sp
279	6	1.4	318	1	LP5A_BACNO	P39907	bacteroides	352	6	1.4	378	1	HAIR_DROVI	P29303	drosophila
280	6	1.4	319	1	PTHB_ECOLI	P56580	escherichia	353	6	1.4	380	1	ELYA_BACAO	P27693	bacillus al
281	6	1.4	319	1	ROAL_MACMU	O28521	macaca mula	354	6	1.4	380	1	ELYA_BACCS	P41362	bacillus ci
282	6	1.4	319	1	ROAL_MOUSE	P49312	mus muscucu	355	6	1.4	381	1	NIFV_KLEPN	P05345	klebsiella
283	6	1.4	319	1	ROAL_RAT	P04256	rattus norv	356	6	1.4	381	1	YU06_CAEEL	O08169	apis mellif
284	6	1.4	321	1	ACCO_DIACA	P31528	dianthus ca	357	6	1.4	382	1	HUGA_APIME	P51888	homo sapien
285	6	1.4	322	1	RM10_YEAST	P36520	saccharomyc	358	6	1.4	382	1	PARG_HUMAN	O40224	lupinus lut
286	6	1.4	322	1	YALE_RHSN	P55546	rhicobium s	359	6	1.4	384	1	GBA1_LOPLU	O75056	homo sapien
287	6	1.4	325	1	CYRK_CITLA	O43317	citullus l	360	6	1.4	384	1	SDC3_HUMAN	O03167	saccharomyc
288	6	1.4	325	1	CYRK_SPIOL	O00834	splinacla ol	361	6	1.4	386	1	YMK7_YEAST	P43693	gallus gall
289	6	1.4	325	1	SBCD_BACSU	P23479	bacillus su	362	6	1.4	387	1	GAT6_CHICK	P124711	tarsus ben
290	6	1.4	327	1	PL11_DROME	P48461	drosophila	363	6	1.4	387	1	INVO_TARBA	O10209	schizosacch
291	6	1.4	328	1	FLIG_AQUAE	O66891	aquifex aeo	364	6	1.4	392	1	YAY1_SCHPO	O59228	bacillus st
292	6	1.4	329	1	EXEB_PSEPU	O05605	pseudomonas	365	6	1.4	393	1	AAT_BACST	O55027	synecococc
293	6	1.4	330	1	FERB_ECOLI	P23877	escherichia	366	6	1.4	393	1	SRPC_SYNP7	O15535	homo sapien
294	6	1.4	330	1	YDEZ_ECOLI	P7651	escherichia	367	6	1.4	394	1	Z193_HUMAN	P14091	homo sapien
295	6	1.4	332	1	ILVC_STRCO	O92565	streptomyce	368	6	1.4	396	1	CATE_HUMAN	CHSA_IPOPU	ipomoea
296	6	1.4	332	1	SRG3_CAEEL	P46572	caenorhabdi	369	6	1.4	396	1	DCOP_SORMA	P78749	sordaria ma
297	6	1.4	334	1	RIRZ_ASFB7	P42492	african swi	370	6	1.4	396	1	DCOP_SORMA	P78749	sordaria ma
298	6	1.4	335	1	YJ89_PYRHO	O57713	pyrococcus	371	6	1.4	397	1	PGK_XANFL	P50314	xanthobacte
299	6	1.4	337	1	HAIR_DROME	P14003	drosophila	372	6	1.4	397	1	TYRB_ECOLI	O31712	bacillus su
300	6	1.4	339	1	KMOS_RAT	P00539	rattus norv	373	6	1.4	397	1	YKXZ_BACSU	P36381	gallus gall
301	6	1.4	340	1	AA2B_CHICK	O13076	gallus gall	374	6	1.4	399	1	YKXZ_BACSU	P36381	gallus gall
302	6	1.4	340	1	UL20_HCMVA	P16758	human cytom	375	6	1.4	400	1	CX48_CHICK	P52193	bos taurus
303	6	1.4	340	1	YDDR_ECOLI	P73108	escherichia	376	6	1.4	400	1	AMPC_PSYIM	O05465	psychobact
304	6	1.4	341	1	MOAA_BACSU	P39757	bacillus su	377	6	1.4	401	1	ARRS_MOUSE	P20443	mus muscucu
305	6	1.4	342	1	KMOS_MSPTS	P10421	myeloprolif	378	6	1.4	403	1	ARRS_MOUSE	P20443	mus muscucu
306	6	1.4	344	1	CHSA_PLANT	P48395	phablicits n	379	6	1.4	403	1	ARRS_MOUSE	P20443	mus muscucu
307	6	1.4	344	1	HOLA_HAEIN	P43747	haemophilus	380	6	1.4	404	1	IMPL_EBYC	P23362	epstein-bar
308	6	1.4	345	1	CEBB_HUMAN	P17676	homo sapien	381	6	1.4	404	1	IMPL_EBYC	P23362	epstein-bar
309	6	1.4	346	1	YAOD_SCHPO	O09681	schizosacch	382	6	1.4	405	1	YLR2_YEAST	P07267	saccharomyc
310	6	1.4	347	1	VP11_WTV	P13094	wound tumor	383	6	1.4	405	1	YLR2_YEAST	P07267	saccharomyc
311	6	1.4	349	1	25A1_PIG	O29599	sus scrofa	384	6	1.4	405	1	YLR2_YEAST	P07267	saccharomyc
312	6	1.4	349	1	OAS1_PIG	O29599	sus scrofa	385	6	1.4	406	1	TRPB_RHIER	P58746	saccharomyc
313	6	1.4	351	1	HRCA_MYCPN	P75351	mycoplasma	386	6	1.4	407	1	TRPB_RHIER	P58746	saccharomyc
314	6	1.4	351	1	RFBX_XANCA	P55295	xanthomonas	387	6	1.4	407	1	TRPB_RHIER	P58746	saccharomyc
315	6	1.4	351	1	YITG_ECOLI	P32151	escherichia	388	6	1.4	411	1	SYT_THIPE	P41256	thiobacilli
316	6	1.4	352	1	CHSC_IPOPU	P48399	ipomoea pur	389	6	1.4	411	1	SYT_THIPE	P41256	thiobacilli
317	6	1.4	352	1	CICE_PSEAE	O87612	pseudomonas	390	6	1.4	411	1	SYT_THIPE	P41256	thiobacilli
318	6	1.4	352	1	CICE_PSEAE	O87612	pseudomonas	391	6	1.4	411	1	SYT_THIPE	P41256	thiobacilli
319	6	1.4	353	1	YMG3_CAEEL	O30847	pseudomonas	392	6	1.4	412	1	YMG3_CAEEL	P32950	candida par
320	6	1.4	353	1	TCPH_TETTH	P54410	tetrahymena	393	6	1.4	412	1	YMG3_CAEEL	P32950	candida par
321	6	1.4	354	1	TFPI_ALCEU	P27137	alcaligenes	394	6	1.4	415	1	EUTH_SALTY	P41796	salmonella
322	6	1.4	356	1	NTRB_RHOCA	P09431	rhodobacter	395	6	1.4	415	1	EUTH_SALTY	P41796	salmonella
323	6	1.4	359	1	AC14_CHICK	P50333	gallus gall	396	6	1.4	415	1	EUTH_SALTY	P41796	salmonella
324	6	1.4	359	1	DHOM_YEAST	P31116	saccharomyc	397	6	1.4	416	1	CRIC_MOUSE	P14211	mus muscucu
325	6	1.4	359	1	GCPE_HELPY	O92110	helicobacte	398	6	1.4	416	1	CRIC_MOUSE	P14211	mus muscucu

399	1.4	417	1	CRTC_HUMAN	P27297	homo sapien	472	1.4	483	1	MORE_CHLTR	O84271	chlamydia t
400	1.4	417	1	DHMH_PARDE	P29894	paracoccus	473	1.4	483	1	XYIB_STAXY	P27155	staphylococ
401	1.4	417	1	Z332_HUMAN	Q9unys	homo sapien	474	1.4	485	1	OPRM_PSEAE	O51487	staphylococ
402	1.4	418	1	CRTC_RABIT	P15253	oryctolagus	475	1.4	485	1	Z165_HUMAN	P49910	homo sapien
403	1.4	419	1	VEGC_HUMAN	P49767	homo sapien	476	1.4	486	1	EXU_DROVI	O24747	drosophila
404	1.4	420	1	EFIA_HALMA	P16018	haloarcuella	477	1.4	487	1	Y442_MYCTU	P42611	mycobacteri
405	1.4	421	1	EFIA_HALMA	P48863	halobacteri	478	1.4	488	1	ASTD_PSEAE	O50174	pseudomonas
406	1.4	423	1	MB12_YEAST	P48873	saccharomyc	479	1.4	489	1	C128_MYCTU	O59572	mycobacteri
407	1.4	423	1	POLS_SINDM	P11359	sinidbis vir	480	1.4	490	1	TMS2_MOUSE	O91498	mus musculu
408	1.4	424	1	SYT_MYCTU	O33191	mycobacteri	481	1.4	496	1	CMAE_DROME	O46054	drosophila
409	1.4	425	1	HE47_CAEEL	Q18212	caenorhabdi	482	1.4	496	1	CMAA_HUMAN	P21941	homo sapien
410	1.4	427	1	FLR1_HUMAN	O75955	homo sapien	483	1.4	496	1	KPR5_YEAST	Q12265	saccharomyc
411	1.4	428	1	EFIA_PYRHO	O59153	pyrococcus	484	1.4	498	1	GATA_EMENT	Q91210	emericella
412	1.4	428	1	FLR1_MOUSE	O09817	mus musculu	485	1.4	499	1	GLPK_SUISO	Q91210	delinococcus
413	1.4	428	1	FLR1_RAT	O921e1	rattus norv	486	1.4	499	1	RRG_BRARE	O91392	brachydanio
414	1.4	428	1	MNTH_MYCTU	O05916	mycobacteri	487	1.4	500	1	YDEM_ECOLI	P38282	escherichia
415	1.4	428	1	PROA_TREPA	P74935	treponema p	488	1.4	501	1	LYSI_CONGL	P38863	corynebacte
416	1.4	429	1	YPAV_ECOLI	P76470	escherichia	489	1.4	501	1	SYK_HELPJ	Q94mp8	helicobacte
417	1.4	430	1	EFIA_PYRMO	P26751	pyrococcus	490	1.4	501	1	SYK_HELPJ	P56126	helicobacte
418	1.4	431	1	NPTX_CHICK	O90932	gallus gall	491	1.4	501	1	TX18_HUMAN	O95935	homo sapien
419	1.4	431	1	APRE_PSEAE	O03025	pseudomonas	492	1.4	504	1	SPAS_MOUSE	O94YV8	mus musculu
420	1.4	432	1	YF10_MYCTU	P71789	mycobacteri	493	1.4	504	1	YH99_YEAST	P38854	saccharomyc
421	1.4	433	1	UROK_BOVIN	O05589	bos taurus	494	1.4	505	1	PR12_MOUSE	O97236	homo sapien
422	1.4	435	1	R114_YEAST	P32841	saccharomyc	495	1.4	505	1	C7C3_ARATH	P3610	mus musculu
423	1.4	435	1	YVBO_BACSU	P37489	baecillus su	496	1.4	507	1	RECN_CAMJE	O64638	arabidopsis
424	1.4	435	1	YVSG_CAEEL	P49191	caenorhabdi	497	1.4	507	1	ASPR_HORVU	P42210	herdenu vul
425	1.4	436	1	EPSE_BURSO	O45111	burkholderi	498	1.4	508	1	GGPD_ANASP	P46992	anabena sp
426	1.4	437	1	PEPG_IACDL	P94869	lactobacilli	499	1.4	509	1	GGPD_STN3	P73411	synecocyst
427	1.4	439	1	Y441_BUCAT	P57316	buchnera ap	500	1.4	509	1	NCAP_SVS	O86435	smilian viru
428	1.4	440	1	LCAT_RABIT	O06774	oryctolagus	501	1.4	509	1	PR12_HUMAN	P49643	homo sapien
429	1.4	441	1	GSA_PROER	O06774	propionibac	502	1.4	509	1	ACHG_XENLA	P05376	xenopus lae
430	1.4	442	1	HPD_DAUCA	O23920	daucus caro	503	1.4	510	1	NOAL_HUMAN	P51513	homo sapien
431	1.4	446	1	DCUB_ECOLI	P14409	escherichia	504	1.4	510	1	VAJD_RHISN	P55504	rhizobium s
432	1.4	447	1	EFIA_PBA	O41011	pisum sativ	505	1.4	511	1	SYS_METTH	O27194	methanobact
433	1.4	447	1	EFIA_SOYBN	P23698	glycine max	506	1.4	513	1	SP12_RABIT	P00187	oryctolagus
434	1.4	449	1	EFIA_VICFA	O24534	vicia faba	507	1.4	515	1	CP12_RABIT	P28708	saccharomyc
435	1.4	449	1	EF11_DAUCA	P29521	daucus caro	508	1.4	518	1	K1K6_YEAST	O18677	caenorhabdi
436	1.4	450	1	MB21_ARATH	O80998	arabidopsis	509	1.4	520	1	DEY1_CAEEL	O10177	schizosacch
437	1.4	450	1	CBPA_SYNP7	P39660	synecococc	510	1.4	521	1	YAVB_SCHPO	G6PD_SYNP7	synecococc
438	1.4	452	1	PTCC_ECOLI	P17334	escherichia	511	1.4	524	1	PE0_LISWE	O01893	listeria we
439	1.4	453	1	YW25_CAEEL	O11095	caenorhabdi	512	1.4	524	1	CLK3_RAT	P15384	rattus norv
440	1.4	454	1	YUW6_CAEEL	P52715	caenorhabdi	513	1.4	525	1	MOTL_CAUCR	O91P66	caulobacter
441	1.4	455	1	DBDR_XENLA	P42290	xenopus lae	514	1.4	526	1	CLK3_MOUSE	P16390	mus musculu
442	1.4	458	1	NIFN_AZOVI	P10336	azotobacter	515	1.4	529	1	MB1_ECOLI	P08098	escherichia
443	1.4	461	1	ATOC_ECOLI	O06065	escherichia	516	1.4	532	1	HUY_HALL17	P29143	halophilic
444	1.4	461	1	CDST_HUMAN	O92803	h phosphati	517	1.4	532	1	VS13_TRYRB	O90592	gallus gall
445	1.4	461	1	EFIA_ARTSA	P02993	artemia sal	518	1.4	533	1	NCB1_BOVIN	O904B5	bos taurus
446	1.4	462	1	CATC_MOUSE	P97821	mus musculu	519	1.4	533	1	GNIB_HUMAN	O01061	homo sapien
447	1.4	462	1	EFIA_BRARE	O92005	brachydanio	520	1.4	534	1	SYEM_YEAST	P48525	saccharomyc
448	1.4	464	1	CDK8_HUMAN	P49336	homo sapien	521	1.4	536	1	IMAI_SCHPO	O14063	schizosacch
449	1.4	465	1	IMP2_DROME	P29681	drosophila	522	1.4	536	1	SYEM_YEAST	P34589	acholeplasm
450	1.4	466	1	SK10_HUMAN	P56693	homo sapien	523	1.4	542	1	ODP2_MCHLA	O48318	halobacteri
451	1.4	466	1	SK10_MOUSE	O04888	mus musculu	524	1.4	544	1	ODP2_PSEAE	Q15722	caenorhabdi
452	1.4	466	1	SK10_RAT	O55170	rattus norv	525	1.4	545	1	SYN_CAEEL	O56363	pseudomonas
453	1.4	466	1	SYN_BUCAT	P57441	buchnera ap	526	1.4	545	1	ODP2_PSEAE	P10723	brugia mala
454	1.4	467	1	BIAR_BOVIN	O91966	bos taurus	527	1.4	548	1	SYN_BRUMA	P18696	emericella
455	1.4	467	1	E2BD_SCHPO	O09924	schizosacch	528	1.4	550	1	PURX_EMENT	O29979	archaeoglob
456	1.4	468	1	BIAR_PIG	O28998	sus scrofa	529	1.4	551	1	SYE_ARCTU	O61712	mus musculu
457	1.4	468	1	PPAD_YEAST	P52290	saccharomyc	530	1.4	552	1	WT01_MOUSE	P51912	mus musculu
458	1.4	468	1	SPSR_BACSU	P37875	baecillus su	531	1.4	552	1	AAAX_MOUSE	Q07231	mus musculu
459	1.4	473	1	BIAR_CANFA	P79148	canis famill	532	1.4	553	1	ZE38_MOUSE	O53728	streptomyc
460	1.4	474	1	BIAR_PELCA	O91626	felis silve	533	1.4	555	1	PLD_STRAT	P54409	tetrahymena
461	1.4	475	1	DLDH_VIBPA	P24695	thiobacilli	534	1.4	556	1	FILE_SALTY	P15928	salmonella
462	1.4	475	1	RP54_THIFE	P08588	homo sapien	535	1.4	559	1	YH12_ARATH	O22757	arabidopsis
463	1.4	477	1	BIAR_HUMAN	P90597	typanosoma	536	1.4	560	1	GNM1_ACICE	P34583	acidothermum
464	1.4	477	1	DLDH_TRYCR	Q06515	arabidopsis	537	1.4	562	1	DP3X_BACSU	P09122	baecillus su
465	1.4	480	1	E6A_ARATH	P47899	macaca mula	538	1.4	563	1	PK66_PLAKU	P21303	plasmidium
466	1.4	481	1	LMRA_STRLN	P46104	streptomyc	539	1.4	565	1	ES10_RAT	P16303	rattus norv
467	1.4	481	1	Y209_ARCFU	O30030	archaeoglob	540	1.4	566	1	MX1D_SHISO	O55293	shigella so
468	1.4	481	1	YEAN_ECOLI	P76552	escherichia	541	1.4					
469	1.4	482	1	PURB_SCHPO	O60105	schizosacch	542	1.4					
470	1.4	482	1				543	1.4					
471	1.4	483	1	FIBA_BOVIN	P02672	bos taurus	544	1.4					

545	6	1.4	567	1	KDGE_HUMAN	P52429	homo sapien	723	1	SSRP_DROME	O05344	drosophila
546	6	1.4	572	1	ND5M_LOCOM	Q36428	Locusta mig	724	1	P85B_BOVIN	P23176	bos taurus
547	6	1.4	575	1	I10R_MOUSE	017577	mus musculus	727	1	IF2M_HUMAN	P46139	homo sapien
548	6	1.4	576	1	PRIM_LEGPN	P71481	legionella	727	1	VP4_RDVA	P22474	rice dwarf
549	6	1.4	576	1	YITO_YEAST	P40568	saccharomyc	727	1	VP4_RDVA	O85436	rice dwarf
550	6	1.4	577	1	ALG8_YEAST	P40351	saccharomyc	728	1	P85B_HUMAN	O00439	homo sapien
551	6	1.4	577	1	STS_RAT	P15589	rattus norv	728	1	AGAL_PEDPE	P43467	pedicococcus
552	6	1.4	579	1	YH24_YEAST	P38880	saccharomyc	733	1	ADDA_MOUSE	O93450	mus musculus
553	6	1.4	580	1	IPAB_SHIDY	O03945	shigella dy	735	1	ADDA_MOUSE	O63078	rattus norv
554	6	1.4	580	1	IPAB_SHIFL	P18011	shigella fl	738	1	PLAP_HUMAN	O93253	homo sapien
555	6	1.4	583	1	SYR_AQUAE	O67068	aquilex aeo	740	1	FAS_PNECA	P29251	p folic aci
556	6	1.4	583	1	YF21_MYCTU	O50566	mycobacteri	741	1	ERF2_PICPI	P23637	pichia pinu
557	6	1.4	584	1	PMEL_CITSI	O04866	citrus sine	742	1	YFA_MOUSE	P23607	mus musculus
558	6	1.4	585	1	GLPD_MYCLE	P53435	mycobacteri	747	1	CLC4_MOUSE	O61418	mus musculus
559	6	1.4	585	1	YAOH_SCHPO	O10096	schizosacch	757	1	CLC4_RAT	P51794	rattus norv
560	6	1.4	590	1	THIC_BACSU	P45740	bacillus su	748	1	HEPA_HSVBC	P52374	bovine hefp
561	6	1.4	592	1	PTI_ALCEU	P23536	alcaligenes	749	1	SMAP_CAEBL	O10560	caenorhabd
562	6	1.4	593	1	Y917_MYCTU	O05909	mycobacteri	756	1	PIDI_HUMAN	P51178	homo sapien
563	6	1.4	594	1	DHBK_LYCES	O04059	lycoperisico	758	1	CLPA_HUMAN	P15716	escherichia
564	6	1.4	594	1	RBJK_DROME	P28159	drosophila	760	1	CLC4_HUMAN	P51794	homo sapien
565	6	1.4	602	1	LEPA_CHLPN	O92814	chlamydia p	767	1	PUBI_SCHPO	O92462	schizosacch
566	6	1.4	602	1	LEPA_CHLTR	O84067	chlamydia t	767	1	YNM4_YEAST	P53866	saccharomyc
567	6	1.4	603	1	LCB2_SCHPO	O09925	schizosacch	770	1	PCR_MOUSE	O61139	mus musculus
568	6	1.4	606	1	GAB_DROME	P25123	drosophila	771	1	CTP6_MYCTU	O10866	mycobacteri
569	6	1.4	609	1	YLO1_YEAST	O06033	saccharomyc	773	1	MSL2_DROME	P50534	drosophila
570	6	1.4	610	1	THIC_RHET	O34291	rhizobium e	775	1	MGMS_YEAST	P32794	saccharomyc
571	6	1.4	612	1	Y525_PSEAE	O51484	pseudomonas	780	1	AFG2_YEAST	O62849	rattus norv
572	6	1.4	612	1	YMYE_YEAST	O03153	saccharomyc	783	1	PCR_RAT	O16549	homo sapien
573	6	1.4	613	1	MLIX_HUMAN	O13585	homo sapien	785	1	PCR_HUMAN	P21010	vacchina vl
574	6	1.4	619	1	IF2_MYCG	P47388	mycoplasma	785	1	VD05_VACCC	P21010	vacchina vl
575	6	1.4	623	1	OGP_PAPAN	P36717	papilo anub1	785	1	VD05_VACCV	P33069	variolia vlr
576	6	1.4	623	1	PNT1_DROME	P51022	drosophila	785	1	VD05_VAVR	O99148	y bifunctio
577	6	1.4	625	1	TBP2_HAEIN	P44991	haemophilus	788	1	PUR2_YARLI	O99148	xanthomonas
578	6	1.4	629	1	ODP2_ECOLI	P06959	escherichia	790	1	QUIA_XANCU	O99148	xanthomonas
579	6	1.4	638	1	OAR1_LYMT	O77408	lymaea sta	794	1	TRK1_LYMT	O76997	lymaea sta
580	6	1.4	639	1	TETM_STRLI	O02652	streptomyce	794	1	Z151_MOUSE	O60821	mus musculus
581	6	1.4	642	1	YB65_SCHPO	O09746	schizosacch	799	1	ZFX1_MOUSE	P17011	mus musculus
582	6	1.4	645	1	IME2_YEAST	P32581	saccharomyc	800	1	ZFX_BOVIN	O62849	bos taurus
583	6	1.4	645	1	SUBE_BACSU	P16396	bacillus su	801	1	EGR3_MOUSE	O61851	mus musculus
584	6	1.4	646	1	PLAP_MOUSE	P27612	mus musculus	801	1	ZFX_HUMAN	P08048	homo sapien
585	6	1.4	646	1	VP40_HSVB	P28936	equine hefp	803	1	Z151_HUMAN	O13105	homo sapien
586	6	1.4	647	1	PLAP_RAT	P54319	rattus norv	805	1	IF2_AQUAE	O67825	aquilex aeo
587	6	1.4	651	1	YHJK_ECOLI	P37649	escherichia	805	1	SUXI_TULGE	O41608	tulipa gesn
588	6	1.4	654	1	TPE2_HUMAN	P15923	homo sapien	806	1	ZFX_HUMAN	P17010	homo sapien
589	6	1.4	656	1	YK79_MYCTU	O10687	mycobacteri	806	1	EGR3_HUMAN	P22607	homo sapien
590	6	1.4	660	1	VG06_BP74	P19060	bacterioph	806	1	LON_RHIME	O69177	rhizobium m
591	6	1.4	661	1	ACSA_COPCI	O13440	coprinus ci	806	1	LON_BROAB	O52605	bruce11a ab
592	6	1.4	665	1	Y176_MERTH	O26218	methanobact	812	1	NAH1_HUMAN	P19634	homo sapien
593	6	1.4	667	1	SIM2_HUMAN	O14190	homo sapien	815	1	PPSA_PYRFU	P42850	pyrococcus
594	6	1.4	670	1	INVI_MAIZE	P49175	zea mays (m	817	1	TRBE_AGRTU	P54910	agrobacteri
595	6	1.4	677	1	FLHA_BACSU	P35620	bacillus su	820	1	GYRA_CIOAB	P94605	clostridium
596	6	1.4	678	1	OGP_HUMAN	O12889	homo sapien	829	1	IF2_HAEIN	P44323	haemophilus
597	6	1.4	678	1	YF48_MYCTU	O10778	mycobacteri	829	1	PKN2_MYXXA	P54736	myxococcus
598	6	1.4	680	1	GAG_SCVLA	P32503	saccharomyc	830	1	PRLR_COLLI	O90373	columbia liv
599	6	1.4	687	1	TRFE_ONCKI	P79815	oncothychnu	830	1	VP3_HUMAN	O13438	h vacuolar
600	6	1.4	690	1	IF2_THEMA	O97243	thermotoga	834	1	MSH5_HUMAN	O43196	homo sapien
601	6	1.4	690	1	TOP1_BACHD	O9ka23	bacillus ha	834	1	GYRA_MYCTU	O07702	mycobacteri
602	6	1.4	692	1	DNK1_STY3	O55154	synecocyst	839	1	ZFX_MOUSE	P17012	mus musculus
603	6	1.4	692	1	REFX_MOUSE	P48371	mus musculus	842	1	EF2_YEAST	P32334	saccharomyc
604	6	1.4	695	1	YD38_YEAST	O05497	saccharomyc	842	1	GYRA_MYCSM	P48361	tyreponema p
605	6	1.4	696	1	Y012_CAEBL	P34668	caenorhabd1	842	1	IF2_TREPA	O83611	schizosacch
606	6	1.4	698	1	YB06_YEAST	P38283	saccharomyc	848	1	TOP1_BORBU	O51768	schizosacch
607	6	1.4	700	1	FLHA_CANCR	O03845	caulobacter	854	1	RA54_SCHPO	P41410	schizosacch
608	6	1.4	701	1	YF63_BUCAT	P57444	buchnera ap	854	1	SYV_THERH	P96142	thermus aqu
609	6	1.4	703	1	UL17_HSV11	P10201	herpes simp	855	1	CNRC_BOVIN	P16586	bos taurus
610	6	1.4	706	1	GY5_NEUCR	O93869	neurospora	860	1	LDLR_HUMAN	P01130	homo sapien
611	6	1.4	712	1	NRD2_ECOLI	P28903	escherichia	863	1	YF67_CAEBL	O09212	caenorhabd1
612	6	1.4	718	1	PNT2_DROME	P51023	drosophila	867	1	EF2_BLAHO	O17152	blastocysti
613	6	1.4	720	1	AGAL_STRMU	P27756	streptococc	868	1	TOP1_HAEIN	P43012	haemophilus
614	6	1.4	721	1	Y781_METJA	O58191	methanococc	868	1	VAV2_MOUSE	O60992	mus musculus
615	6	1.4	722	1	P85B_MOUSE	O08908	mus musculus	870	1	Y563_HUMAN	O60309	homo sapien
616	6	1.4	722	1	P85B_RAT	O63788	rattus norv	872	1	SCDI_SCHPO	P40995	schizosacch
617	6	1.4	723	1	REFX2_HUMAN	P48378	homo sapien	873	1	FIMC_BORPE	P33410	bordeletella

691	6	1.4	874	1	SLAP_BACLI	P49052 bacillus 11	764	6	1.4	1202	1	NOS3_HUMAN	P29474 homo sapien
692	6	1.4	876	1	TOPI_VIBCH	Q9krb2 vibrio chol	765	6	1.4	1204	1	NOS3_PIG	Q28969 sus scrofa
693	6	1.4	877	1	DPO1_LACIC	Q32801 lactococcus	766	6	1.4	1205	1	YAW2_SCHPO	Q10178 schizosacch
694	6	1.4	878	1	YAW2_HUMAN	P52735 homo sapien	767	6	1.4	1209	1	THR_DROME	P42286 drosophila
695	6	1.4	880	1	ENV_STIML	P11267 simlan immu	768	6	1.4	1215	1	BGCN_DROME	Q9w112 drosophila
696	6	1.4	881	1	ENV_STIMK	P05884 simlan immu	769	6	1.4	1220	1	ATCP_HUMAN	P20020 homo sapien
697	6	1.4	889	1	ENV_SIVSP	P19503 simlan immu	770	6	1.4	1220	1	ATCP_PIG	Q23200 sus scrofa
698	6	1.4	890	1	IF2_ECOLI	P02995 escherichia	771	6	1.4	1220	1	ATCP_RABIT	Q00804 oryctolagus
699	6	1.4	892	1	YCF2_MESVI	Q9mpu8 mesostigma	772	6	1.4	1224	1	RBP2_YEAST	P08818 saccharomyc
700	6	1.4	892	1	AAC1_HUMAN	P12814 homo sapien	773	6	1.4	1228	1	ALAS_ARATH	Q98993 arabidopsis
701	6	1.4	892	1	AAC1_RAT	Q921p2 rattus norv	774	6	1.4	1229	1	KPB2_FUGRU	Q9w0r1 fugu rubrip
702	6	1.4	892	1	ATX7_HUMAN	O15265 homo sapien	775	6	1.4	1238	1	DPOG_MOUSE	P54099 mus musculu
703	6	1.4	892	1	IF2_SALTY	Q92f31 salmonella	776	6	1.4	1245	1	POIS_SINDY	P03316 sindbis vir
704	6	1.4	893	1	AAC1_CHICK	P05094 gallus gall	777	6	1.4	1246	1	YAW2_CAEEL	P34504 caenorhabdi
705	6	1.4	894	1	WPRB_BACSU	P54233 bacillus su	778	6	1.4	1248	1	CYAB_RAT	P40146 rattus norv
706	6	1.4	896	1	BOSS_DROME	P22815 drosophila	779	6	1.4	1249	1	CYAB_MOUSE	P97990 mus musculu
707	6	1.4	896	1	IF2_KLROX	Q92f28 klebsiella	780	6	1.4	1251	1	CYAB_HUMAN	P40145 homo sapien
708	6	1.4	897	1	IF2_ENTCL	Q92f25 enterobacte	781	6	1.4	1261	1	YGN3_YEAST	P53125 saccharomyc
709	6	1.4	898	1	FAS2_SCHAM	P22648 schistocerc	782	6	1.4	1265	1	DYNA_DROME	P13496 drosophila
710	6	1.4	898	1	NIA_BETVE	P27783 betula verr	783	6	1.4	1270	1	DYNA_HUMAN	Q14203 homo sapien
711	6	1.4	900	1	IF2_MYCTU	P71613 mycobacteri	784	6	1.4	1273	1	GYRA_MYCLE	O57532 mycobacteri
712	6	1.4	900	1	VGIB_HSVIF	P06436 herpes simp	785	6	1.4	1310	1	ICP4_VZVD	P09310 varicella-z
713	6	1.4	904	1	VGIB_HSV11	P10211 herpes simp	786	6	1.4	1323	1	NME4_MOUSE	Q03391 mus musculu
714	6	1.4	904	1	VGIB_HSVIP	P08665 herpes simp	787	6	1.4	1336	1	MAM1_SCHPO	P78966 schizosacch
715	6	1.4	915	1	SMOO_HUMAN	P53814 homo sapien	788	6	1.4	1342	1	XDH_DROPS	P22811 drosophila
716	6	1.4	916	1	RTJK_DROME	P21328 drosophila	789	6	1.4	1357	1	RPOB_PSEPU	P19175 pseudomonas
717	6	1.4	917	1	IF2_PROVU	Q92f22 proteus vul	790	6	1.4	1377	1	NEO1_RAT	P97603 rattus norv
718	6	1.4	920	1	MMI2_MYCTU	P96289 mycobacteri	791	6	1.4	1391	1	NI57_YEAST	P40064 saccharomyc
719	6	1.4	921	1	NAC2_HUMAN	Q9upr5 homo sapien	792	6	1.4	1396	1	ITR2_DROME	P12080 drosophila
720	6	1.4	921	1	NAC2_RAT	P48768 rattus norv	793	6	1.4	1433	1	REST_CHICK	Q42184 gallus gall
721	6	1.4	929	1	KDG2_RAT	O08560 bacillus su	794	6	1.4	1434	1	NOS1_HUMAN	P29475 homo sapien
722	6	1.4	937	1	ODO1_BACSU	P23129 bacillus su	795	6	1.4	1435	1	NOS1_RABIT	O19132 oryctolagus
723	6	1.4	943	1	NFC1_HUMAN	O95644 homo sapien	796	6	1.4	1456	1	RPO_PVX	P09395 potato viru
724	6	1.4	946	1	YIA2_YEAST	P40559 saccharomyc	797	6	1.4	1456	1	RPO_PVXCP	P22591 potato viru
725	6	1.4	952	1	YH19_RHOCA	O02998 rhodobacter	798	6	1.4	1456	1	RPO_PVX3	P17779 potato viru
726	6	1.4	954	1	M3KA_HUMAN	Q02779 homo sapien	799	6	1.4	1461	1	NEO1_HUMAN	Q92859 homo sapien
727	6	1.4	959	1	SYAC_SCHPO	O13314 schizosacch	800	6	1.4	1476	1	AZM1_MOUSE	P28665 mus musculu
728	6	1.4	960	1	CAP1_SORBI	P29195 sorghum blc	801	6	1.4	1493	1	NEO1_MOUSE	P97798 mus musculu
729	6	1.4	960	1	CAP2_MESCR	P16097 mesembryant	802	6	1.4	1589	1	PHP_DROME	P39769 drosophila
730	6	1.4	961	1	TSP4_HUMAN	P35443 homo sapien	803	6	1.4	1616	1	GLSE_MAIZE	P23225 zea mays (m
731	6	1.4	962	1	YAE_SCHPO	O09731 schizosacch	804	6	1.4	1788	1	POIN_SOUY3	Q04545 southampton
732	6	1.4	966	1	MMIB_MYCTU	O53653 mycobacteri	805	6	1.4	1938	1	MYSD_CAEEL	P02567 caenorhabdi
733	6	1.4	966	1	ST10_MOUSE	O55098 mus musculu	806	6	1.4	1966	1	CCAF_HUMAN	Q06040 homo sapien
734	6	1.4	967	1	CAP1_SOYBN	Q02909 glycine max	807	6	1.4	1972	1	P531_HUMAN	O12888 homo sapien
735	6	1.4	968	1	ST10_HUMAN	O94804 homo sapien	808	6	1.4	2035	1	HFC1_HUMAN	P51610 homo sapien
736	6	1.4	977	1	EPA2_MOUSE	Q03405 mus musculu	809	6	1.4	2109	1	RRLP_VSVJH	P16315 vesicular s
737	6	1.4	980	1	YRS5_CAEEL	O09349 caenorhabdi	810	6	1.4	2109	1	RRLP_VSVYO	P16379 vesicular s
738	6	1.4	982	1	YSG6_CAEEL	O09365 caenorhabdi	811	6	1.4	2109	1	RRLP_VSVST	P03523 vesicular s
739	6	1.4	986	1	GUN2_CLOSR	P23659 clostridium	812	6	1.4	2142	1	BAT2_HUMAN	P46634 homo sapien
740	6	1.4	994	1	SVY_XYLFA	O9ph12 xyella fas	813	6	1.4	2208	1	POIN_MANCY	O69014 manchester
741	6	1.4	1006	1	EPB6_HUMAN	O15197 homo sapien	814	6	1.4	2209	1	Y166_HUMAN	P07259 saccharomyc
742	6	1.4	1030	1	Y075_MYCPN	P75556 mycoplasma	815	6	1.4	2214	1	PYR1_YEAST	O09794 drosophila
743	6	1.4	1036	1	P200_MYCPN	P75211 mycoplasma	816	6	1.4	2244	1	PYR1_SCHPO	O09794 drosophila
744	6	1.4	1039	1	MSI1_DROME	P50535 drosophila	817	6	1.4	2349	1	TPR_HUMAN	P12270 homo sapien
745	6	1.4	1048	1	AGOI_ARATH	O04379 arabidopsis	818	6	1.4	2353	1	CCAH_HUMAN	O95180 homo sapien
746	6	1.4	1057	1	SPS1_CITON	O22600 citius unsh	819	6	1.4	2376	1	Y1W9_YEAST	P40468 saccharomyc
747	6	1.4	1068	1	SPS_MAIZE	P31927 zea mays (m	820	6	1.4	2453	1	NCR1_MOUSE	O60974 mus musculu
748	6	1.4	1079	1	MYIA_MOUSE	P46735 mus musculu	821	6	1.4	2483	1	PCX_DROME	P18490 drosophila
749	6	1.4	1085	1	CUT7_SCHPO	P24339 schizosacch	822	6	1.4	2524	1	NORC_XENLA	P21783 xenopus lae
750	6	1.4	1101	1	FURL_DROME	P30430 drosophila	823	6	1.4	2531	1	NTC1_MOUSE	Q00708 mus musculu
751	6	1.4	1121	1	WDR6_HUMAN	Q9nmw5 homo sapien	824	6	1.4	2531	1	NTC1_RAT	Q00708 mus musculu
752	6	1.4	1141	1	CN3A_HUMAN	Q14432 homo sapien	825	6	1.4	2813	1	VWE_CANFA	Q28295 canis famli
753	6	1.4	1147	1	NRDC_HUMAN	O43847 homo sapien	826	6	1.4	2842	1	APC_CANFA	P70478 rattus norv
754	6	1.4	1151	1	ITAI_HUMAN	P56199 homo sapien	827	6	1.4	2845	1	APC_RAT	O06131 mus musculu
755	6	1.4	1153	1	ATC2_YEAST	P38929 saccharomyc	828	6	1.4	3061	1	POIG_PVYHU	Q02963 p genome po
756	6	1.4	1173	1	ATC2_HUMAN	O16720 homo sapien	829	6	1.4	3063	1	POIG_PVYN	P18247 p genome po
757	6	1.4	1176	1	ATCP_RAT	P15505 rattus norv	830	6	1.4	3068	1	POIG_PEMVC	Q001500 p genome po
758	6	1.4	1180	1	ITAI_RAT	P18614 rattus norv	831	6	1.4	3110	1	HD_RAT	P51111 rattus norv
759	6	1.4	1182	1	RTP2_TRYBG	P15594 trypanosoma	832	6	1.4	3119	1	HD_MOUSE	P42858 mus musculu
760	6	1.4	1185	1	MAPX_DROME	P23326 drosophila	833	6	1.4	3144	1	POIG_TUWYO	P42858 mus musculu
761	6	1.4	1193	1	RPOB_BACSU	P37870 bacillus su	834	6	1.4	3163	1	POIG_TUWVJ	O02597 t genome po
762	6	1.4	1198	1	ATCO_HUMAN	Q01814 homo sapien	835	6	1.4	3164	1	POIG_TUWVJ	P89509 t genome po
763	6	1.4	1198	1	ATCO_RAT	P11506 rattus norv	836	6	1.4	3224	1	RBP2_HUMAN	P49792 homo sapien

837	6	1.4	3358	1	PGCV_MOUSE	062059	mus musculus
838	6	1.4	3412	1	POLG_TBESVS	P07720	t genome po
839	6	1.4	4196	1	DYHC_SCHPO	013290	schizosacch
840	6	1.4	4273	1	PKSM_BACSU	P40872	bacillus su
841	6	1.4	4427	1	PKSL_BACSU	005470	bacillus su
842	6	1.4	4447	1	PKSK_BACSU	P40803	bacillus su
843	6	1.4	4550	1	FATH_HUMAN	Q14517	homo sapien

ALIGNMENTS

```

RESULT 1
ID      NUSM_BRALA      STANDARD;      PRT;      598 AA.
AC      079422;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DE      NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3)
GN      ND5 OR NAD5 OR NADH5.
OS      Branchiostoma lanceolatum (Common Lancelet) (Amphioxus).
OC      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC      Branchiostoma.
OX      NCBI_TaxID=7740;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98292550; PubMed=9628930;
RA      Spruyt N., Delarbre C., Gacheil G., Laudet V.;
RT      Complete sequence of the amphioxus (Branchiostoma lanceolatum)
RT      mitochondrial genome: relations to vertebrates.;
RL      Nucleic Acids Res. 26:3279-3285(1998).
CC      -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Y16474; CAA76257.1; -
CC      DR      InterPro: IPR001516; -
CC      DR      InterPro: IPR001750; -
CC      DR      Pfam: PF00361; oxidored_q1; 1.
CC      DR      Pfam: PF00662; oxidored_q1_N; 1.
CC      KW      Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC      SQ      SEQUENCE 598 AA; 66097 MW; EDA986E0B8B1C565 CRC64;

Query Match      1.88; Score 8; DB 1; Length 598;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      46 LLVSAES 53
DB      122 LLVSAES 129

RESULT 2
ID      NUSM_BRALA      STANDARD;      PRT;      599 AA.
AC      047430;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DE      NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3)
GN      ND5 OR NAD5 OR NADH5.
OS      Branchiostoma floridae (Florida Lancelet) (Amphioxus).
OC      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

```

```

OC      Branchiostoma.
OX      NCBI_TaxID=7739;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99261652; PubMed=10331267;
RA      Boore J.L., Daehler L.L., Brown W.M.;
RT      Complete sequence, gene arrangement, and genetic code of
RT      mitochondrial DNA of the cephalochordate Branchiostoma floridae
RT      (Amphioxus).;
RL      Mol. Biol. Evol. 16:410-418(1999).
CC      -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF098298; AAB87999.1; -
CC      DR      InterPro: IPR001516; -
CC      DR      InterPro: IPR001750; -
CC      DR      Pfam: PF00361; oxidored_q1; 1.
CC      DR      Pfam: PF00662; oxidored_q1_N; 1.
CC      KW      Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC      SQ      SEQUENCE 599 AA; 66140 MW; 9E04167704FA2990 CRC64;

Query Match      1.88; Score 8; DB 1; Length 599;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      46 LLVSAES 53
DB      122 LLVSAES 129

RESULT 3
ID      V120_EBV      STANDARD;      PRT;      1239 AA.
AC      P03189;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      CAPSID ASSEMBLY PROTEIN BOLFI.
GN      BOLFI.
OS      Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Gammaherpesvirinae; Lymphocryptovirus.
OX      NCBI_TaxID=10377;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=84270667; PubMed=6087149;
RA      Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA      Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA      Tufnell P.S., Barrett B.G.;
RT      "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RT      Nature 310:207-211(1984).
RL      Nature 310:207-211(1984).
CC      -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
CC      EBV-1 23, EBV BOLFI, VZV 21, HVS-1 63, AND HCMV UL47.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; V01555; CAA26841.1; -
CC      DR      PIR; A03752; Q0BE10.
CC      DR      PIR; S32995; S32995.

```

KW Capsid assembly. 132748 MW; 6C5DBRC55F2F729 CRC64;
SQ SEQUENCE 1239 AA: 132748 MW; 6C5DBRC55F2F729 CRC64;
Query Match 1.8%; Score 8; DB 1; Length 1239;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 322 RRRLVPA 329
DB 874 RRRLVPA 881
RESULT 4
AKH2.LOCMI STANDARD: PRT: 61 AA.
ID AKH2.LOCMI STANDARD: PRT: 61 AA.
AC P08379; P10617.
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ADIPOKINETIC HORMONE II PRECURSOR (CONTAINS: ADIPOKINETIC HORMONE II
(AKH-II); ADIPOKINETIC HORMONE PRECURSOR-RELATED PEPTIDE BETA CHAIN
(APRP-BETA) (6 KDA DIMERIC PEPTIDE B)).
DE Locusta migratoria (Migratory locust).
OS Locusta migratoria (Migratory locust).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridomorpha; Acridoidea; Acrididae; Locusta.
OX NCBI_TaxID=7004;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Corpora cardiaca;
RC MEDLINE=86032738; PubMed=7559443;
RA Bogerd J., Koolman F.P., Pijnenburg M.A.P., Hekking L.H.,
RA Oudejans R.C., Vander Horst D.J.;
RT "Molecular cloning of three distinct cDNAs, each encoding a different
RT adipokinet hormone precursor, of the migratory locust, Locusta
RT migratoria. Differential expression of the distinct adipokinet
RT hormone precursor genes during flight activity.";
RT J. Biol. Chem. 270:23038-23043(1995).
RN [2]
RP SEQUENCE OF 23-30.
RX MEDLINE=86050918; PubMed=4063072;
RA Siebert K., Morgan P., Mordue W.;
RT "Primary structures of locust adipokinet hormones II.";
RT Biol. Chem. Hoppe-Seyler 366:723-727(1985).
RN [3]
RP SEQUENCE OF 23-30.
RX MEDLINE=86130555; PubMed=3947348;
RA Gaede G., Goldsworthy G.J., Schaffer M.H., Cook J.C.,
RA Rinehart K.L. Jr.;
RT "Sequence analyses of adipokinet hormones II from corpora cardiaca
RT of Schistocerca nitans, Schistocerca gregaria, and Locusta migratoria
RT by fast atom bombardment mass spectrometry.";
RT Biochem. Biophys. Res. Commun. 134:723-730(1986).
RN [4]
RP SEQUENCE OF 34-61.
RX MEDLINE=89276392; PubMed=2731552;
RA Hietter H., Luu B., Goltzene F., Zachary D., Hoffmann J.A.,
RA van Dorsselaer A.;
RT "Isolation and structure of two novel 6-kDa dimeric peptides from the
RT corpora cardiaca of the insect locusta migratoria. Molecular mass
RT determination by mass spectrometry.";
RT Eur. J. Biochem. 182:77-84(1989).
RN [5]
RP FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
RN CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
RN DIACYLGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
RN MUSCLES TO USE THESE DIACYLGLYCERIDES AS AN ENERGY SOURCE.
RN -1- SUBUNIT: ADIPOKINETIC HORMONE PRECURSOR-RELATED PEPTIDE (APRP)
RN CAN FORM THREE TYPE OF DISULFIDE-BOND DIMERS: P1 (ALPHA-ALPHA),
RN P2 (ALPHA-BETA), AND P3 (BETA-BETA).
RN -1- SIMILARITY: BELONGS TO THE AKH / HRH / RPCH FAMILY.
RN THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: X86800; CAA60495.1;
CC DR PIR: B25204; AKIOL2.
CC DR PIR: B24241; B24241.
CC DR InterPro: IPR002047;
CC DR PROSITE: PS00256; AKH; 1.
CC KW Neuropeptide; Amidation; Flight; Cleavage on pair of basic residues;
CC Signal.
FT SIGNAL 1 22
FT CHAIN 23 61
FT PEPTIDE 23 30
FT PEPTIDE 34 61
FT FT
FT MOD_RES 23 23
FT MOD_RES 30 30
FT DISULFID 59 59
FT SEQUENCE 61 AA; 6588 MW; 3436B59C3712A046 CRC64;
Query Match 1.6%; Score 7; DB 1; Length 61;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 37 TLVLVVA 43
DB 8 TLVLVVA 14
RESULT 5
RL22.PRVAP STANDARD: PRT: 79 AA.
ID RL22.PRVAP STANDARD: PRT: 79 AA.
AC 044161;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L22 (FRAGMENT).
GN RPLV OR RPL22.
OS Prunus armeniaca phytoplasma.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Acholeplasmataceae; Phytoplasma.
OX NCBI_TaxID=36589;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94350802; PubMed=8071198;
RA Gundersen D.E., Lee I.M., Rehner S.A., Davis R.E., Kingsbury D.T.;
RT "Phylogeny of mycoplasma-like organisms (phytoplasmas): a basis for
RT their classification.";
RT J. Bacteriol. 176:5244-5254(1994).
RN [2]
RP FUNCTION: THIS PROTEIN BINDS SPECIFICALLY TO 23S RRNA: ITS BINDING
RP IS STIMULATED BY OTHER RIBOSOMAL PROTEINS, E.G., L4, L17, AND L20.
RP IT IS IMPORTANT DURING THE EARLY STAGES OF 50S RECONSTITUTION
RP (BY SIMILARITY).
RN -1- SIMILARITY: BELONGS TO THE L22P FAMILY OF RIBOSOMAL PROTEINS.
RN THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN EMBL: L26994; AA83935.1;
RN DR HSSP: P48286; 1BXE.
RN DR Pfam: PF00237; Ribosomal_L22; 1.
RN DR PROSITE: PS00464; RIBOSOMAL_L22; PARTIAL.


```

NUC_SHIFL
ID NUC_SHIFL STANDARD: PRT: 174 AA.
AC P29769;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE MICROCOCAL NUCLEASE PRECURSOR (EC 3.1.31.1).
GN NUC.
OS Shigella flexneri.
OC Plasmid IncM psa.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219997; PubMed=1560781;
RA Close S.M., Kado C.I.;
RT "A gene near the plasmid psa origin of replication encodes a
RT nuclease."
RT Mol. Microbiol. 6:521-527(1992).
CC -1- SIMILARITY: BELONGS TO THE THERMONUCLEASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: U30471; AAA75246.1; -
CC PIR: S20541; NCEBPX.
CC HSSP: P00644; ISNO.
CC InterPro: IPR000080; -
CC InterPro: IPR002071; -
CC Pfam: PF00565; Snaase_1; 1.
CC PROSITE: PS01123; TNase_1; 1.
CC PROSITE: PS01284; TNase_2; 1.
CC KW Hydrolyase; Nuclease; Endonuclease; Signal; Plasmid.
FT SIGNAL 1 23
FT CHAIN 24 174
FT ACT_SITE 52 52 MICROCOCAL NUCLEASE.
FT ACT_SITE 60 60 BY SIMILARITY.
FT ACT_SITE 94 94 BY SIMILARITY.
SQ SEQUENCE 174 AA; 19689 MW; 982B37729E4F30D7 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 174;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 AA/VLIV 227
DB 13 AA/VLIV 19

RESULT 9
RIMM_MYCLE STANDARD: PRT: 179 AA.
AC 033016;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROBABLE 16S RNA PROCESSING PROTEIN RIMM.
GN RIMM OR MICB250.34.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX Seeger K.J., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- FUNCTION: ESSENTIAL FOR EFFICIENT PROCESSING OF 16S RNA. PROBABLY
CC PART OF THE 30S SUBUNIT PRIOR TO OR DURING THE FINAL STEP IN THE
CC PROCESSING OF 16S FREE 30S RIBOSOMAL SUBUNITS. IT COULD BE SOME
CC ACCESSORY PROTEIN NEEDED FOR EFFICIENT ASSEMBLY OF THE 30S
CC SUBUNIT. RIMM IS NEEDED IN A STEP PRIOR TO RBFA DURING THE
CC MATURATION OF 16S RNA. HAS AFFINITY FOR FREE RIBOSOMAL 30S
CC SUBUNITS BUT NOT FOR 70S RIBOSOMES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RIMM FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: Z97369; CAB10628.1; -
CC InterPro: IPR002676; -
CC Pfam: PF01782; Rimm; 1.
CC RNA processing.
SQ SEQUENCE 179 AA; 19295 MW; F95FFD5C3735930E CRC64;

Query Match 1.6%; Score 7; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 RCDGSEV 124
DB 143 RCDGSEV 149

RESULT 10
YDEQ_BACSU STANDARD: PRT: 197 AA.
AC P96674;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE NAD(P)H OXIDOREDUCTASE YDEQ (EC 1.6.99.-).
GN YDEQ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kasahara Y., Nakai S., Lee S., Sadate Y., Ogasawara N.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE NAD(P)H DEHYDROGENASE (QUINONE) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: AB001488; BAA19364.1; -
CC EMBL: Z99106; CAB12337.1; -
CC Subtilisin; Bg12144; YdeQ.
KW Hypothetical protein; Oxidoreductase.
SQ SEQUENCE 197 AA; 22831 MW; B6BFE72C4AFBFC6E CRC64;

Query Match 1.6%; Score 7; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 36 KTVLVV 42
 Db 5 KTVLVV 11

RESULT 11
 Y027_CAEEL STANDARD; PRT; 243 AA.
 AC P34677;
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE HYPOHETICAL 26.5 KDA PROTEIN ZK688.7 IN CHROMOSOME III.
 GN ZK688.7.
 CC Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Lattelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Shonkeen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "7.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Mature 368332-38(1994).
 CC -1- SIMILARITY: WITH C1365.2.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L16621; AAA28225.1; -
 CC PIR: S44914; S44914.
 CC WormPep: ZK688.7; CE00465.
 CC InterPro: IPR000535; -
 CC Pfam: PF00635; MSP_domain; 1.
 CC Hypothetical protein.
 CC KW
 CC SEQUENCE 243 AA; 26480 MW; 27243B16A74B1C7 CRC64;
 SO
 Query Match 1.6%; Score 7; DB 1; Length 243;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 131 TTRNTVC 137
 Db 142 TTRNTVC 148
 RESULT 12
 MTDH_UROFA STANDARD; PRT; 256 AA.
 AC C00058;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PROBABLE NADP-DEPENDENT MANNITOL DEHYDROGENASE (EC 1.1.1.138) (MTDH)
 DE (MANNITOL 2-DEHYDROGENASE [NADP+]) (PLANTA-INDUCED RUST PROTEIN 8).

GN PIG8.
 OS Uromyces fabae.
 OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Uredinales;
 OC Pucciniaceae; Uromyces.
 OX NCBI_TaxID=55588;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-12; TISSUE-Haustorium;
 RX MEDLINE=97294919; PubMed=9150592;
 RA Hahn M., Mengden K.;
 RT "Characterization of in planta-induced rust genes isolated from a
 RT haustorium-specific cDNA library."
 RL Mol. Plant Microbe Interact. 10:427-437(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN HEXITOL OR PENITOL METABOLISM.
 CC -1- CATALYTIC ACTIVITY: D-MANNITOL + NADP(+) = D-FRUCTOSE + NADPH.
 CC -1- DEVELOPMENTAL STAGE: HAUSTORIA AND RUST-INFECTED LEAVES.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U81790; AAB39878.1; -
 CC InterPro: IPR002198; -
 CC InterPro: IPR002347; -
 CC Pfam: PF00106; adh_short; 1.
 CC Pfam: PF00678; adh_short_C2; 1.
 CC PRINTS: PR00080; SDRFAMILY.
 CC PRINTS: PR00081; GDRBDH.
 CC PROSITE: PS00061; ADH_SHORT; 1.
 CC KW
 CC OXIDOREDUCTASE; NADP.
 CC FT NP_BIND 15 37 NADP (BY SIMILARITY).
 CC ACT_SITE 163 163 BY SIMILARITY.
 CC FT
 CC SEQUENCE 256 AA; 27842 MW; 218677F0010364D5 CRC64;
 SO
 Query Match 1.6%; Score 7; DB 1; Length 256;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 300 VNMLSPG 306
 Db 188 VNMLSPG 194
 RESULT 13
 ARCC_ECOLI STANDARD; PRT; 297 AA.
 AC P37306; P77419;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CARBAMATE KINASE (EC 2.7.2.2).
 GN ARCC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Natch A., Oefner P., Schramm S., Davis R.W.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 203-297 FROM N.A.
RC MEDLINE=89123018; PubMed=2644189;
RA Watanabe W., Sampei G., Alba A., Mizobuchi K.;
RT "Identification and sequence analysis of Escherichia coli pure and
RT purk genes encoding 5'-phosphoribosyl-5-amino-4-imidazole carboxylase
RT for de novo purine biosynthesis."
RL J. Bacteriol. 171:198-204(1989).
RN [4]
RP SEQUENCE OF 203-297 FROM N.A.
RC STRAIN-K12;
RC MEDLINE=89123019; PubMed=2464576;
RA Tiedeman A.A., Keyhani J., Kamholz J., Daum H.A. II, Gots J.S.,
RA Smith J.M.;
RT "Nucleotide sequence analysis of the purk operon encoding 5'-
RT phosphoribosyl-5-aminoimidazole carboxylase of Escherichia coli
RT K-12."
RL J. Bacteriol. 171:205-212(1989).
RN [5]
RP IDENTIFICATION.
RC MEDLINE=95026987; PubMed=7940673;
RA Borodovsky M., Koonin E.V., Rudd K.E.;
RT "New genes in old sequence: a strategy for finding genes in the
RT bacterial genome."
RL Trends Biochem. Sci. 19:309-313(1994).
RN [6]
RP IDENTIFICATION.
RC MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
RT bacterial genome."
RL Nucleic Acids Res. 22:4756-4767(1994).
RN [7]
RP IDENTIFICATION.
RC MEDLINE=95004589; PubMed=7920643;
RA Robison K., Gilbert W., Church G.M.;
RT "Large scale bacterial gene discovery by similarity search."
RL Nat. Genet. 7:205-214(1994).
CC -1- CATALYTIC ACTIVITY: ATP + NH(3) + CO(2) = ADP + CARBAMOYL
CC PHOSPHATE.
CC -1- PATHWAY: THIRD STEP IN ARGININE DEGRADATION VIA ARGININE
CC DETIMINASE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CARBAMATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000158; AAC73623.1; -
DR EMBL; U82664; AAB40273.1; -
DR EMBL; M19657; -; NOT_ANNOTATED_CDS.
DR EMBL; X12982; -; NOT_ANNOTATED_CDS.
DR Ecogene; EG12384; ypcf.
KM Transferase, kinase, Arginine metabolism.
FT CONFLICT 279 297 ALSRIETLAGEACTICISL -> RYRELKRRRAKRGPFVR
FT CSRRH (IN REF. 4).
FT SEQUENCE 297 AA; 31644 MW; 0169371E119AE832 CRC64;

```

Query Match 1.6%; Score 7; DB 1; Length 297;

Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 ICSGGG 253
Db 180 ICSGGG 186

```

RESULT 14
METE_METHH STANDARD; PRT; 308 AA.
ID METE_METHH
AC 026869;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE PROBABLE METHYLCOBALAMIN: HOMOCYSTEINE METHYLTRANSFERASE (EC 2.1.1.-)
DE (METHIONINE SYNTHASE).
DE METE OR MTH775.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Potlier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Carnuso A., Bush D., Sater H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reece J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM
CC METHYLCOBALAMIN AND METHYLCOBINAMIDE TO HOMOCYSTEINE RESULTING IN
CC METHIONINE FORMATION (BY SIMILARITY).
CC -1- COFACTOR: ZINC; BINDS ONE MOLE PER SUBUNIT (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000856; AAB85278.1; -
DR InterPro; IPR002629; -
DR Pfam; PF01717; Methionine_synth; 1.
KM Transferrase; Methyltransferase; Methionine biosynthesis; zinc.
FT METAL 201 201 ZINC (BY SIMILARITY).
FT METAL 203 203 ZINC (BY SIMILARITY).
FT METAL 285 285 ZINC (BY SIMILARITY).
FT METAL 285 285 ZINC (BY SIMILARITY).
FT SEQUENCE 308 AA; 33704 MW; 0157F63D21C80F4F CRC64;

```

Query Match 1.6%; Score 7; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 407 ETLGERL 413
Db 17 ETLGERL 23

RESULT 15
MER4_STRLI

```

ID  MER4_STRL1  STANDARD:  PRT:  319 AA.
AC  P30344;
DT  01-APR-1993 (Rel. 25, Created)
DT  01-APR-1993 (Rel. 25, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  MERCURY RESISTANCE PROBABLE HG TRANSPORT PROTEIN.
OS  Streptomyces lividans.
OC  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC  Streptomycetales; Streptomycetaceae; Streptomycetes.
OX  NCBI_TaxID=1916;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=66 / 1326;
RX  MEDLINE=93156687; PubMed=1494353;
RT  Sedlmeier R., Altenbuchner J.;
RT  "Cloning and DNA sequence analysis of the mercury resistance genes of
CC  Streptomyces lividans."
CC  Mol. Gen. Genet. 236:76-85(1992).
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL: X65467; CAA46465.1; -
DR  PIR: S23613; S23613
DR  PIR: S30173; S30173.
DR  HSSP: P04002; IWFA.
KM  Mercury; Mercuric resistance.
FT  METAL 298 299 HG(2+) (POTENTIAL).
FT  METAL 318 318 HG(2+) (POTENTIAL).
FT  METAL 319 319 HG(2+) (POTENTIAL).
SQ  SEQUENCE 319 AA; 31552 MW; 031495174FD15A14 CRC64;

Query Match 1.68; Score 7; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 AAVLLLV 49
    |||||
Db 174 AAVLLLV 180

RESULT 16
HOLB_PSEAE STANDARD: PRT: 328 AA.
AC PS2024;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA POLYMERASE III, DELTA' SUBUNIT (EC 2.7.7.7).
GN HOLB OR PA2961.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=96125227; PubMed=8550441;
RX Alm R.A., Bodero A.J., Free P.D., Mattick J.S.;
RT "Identification of a novel gene, pilZ, essential for type 4 fimbrial
RT biogenesis in Pseudomonas aeruginosa."
RT J. Bacteriol. 178:46-53(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;

```

```

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huynh W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Lardig R., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RT Nature 406:959-964(2000).
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLYMER COMPLEX. POLYMER ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR  EMBL: L42622; AAA93518.1; -
DR  EMBL: AE004722; AAG06349.1; -
DR  HSSP: P28631; 1A5T.
KM  Transferrase; DNA-directed DNA polymerase; DNA replication.
FT  CONFLICT 16 30 GGRQNHAVLLVGR -> RPRPRRLSLAS (IN
FT  CONFLICT 48 52 QGRPA -> TRAC (IN REF. 1).
FT  CONFLICT 176 177 RA -> PG (IN REF. 1).
FT  CONFLICT 208 209 VR -> AG (IN REF. 1).
FT  CONFLICT 238 251 SVPLPLFDWFCDW -> RVPALALRLVRL (IN REF.
FT  CONFLICT 285 286 QA -> HR (IN REF. 1).
FT  CONFLICT 322 322 A -> G (IN REF. 1).
SQ  SEQUENCE 328 AA; 35746 MW; 55DAA4F223C1464D CRC64;

Query Match 1.68; Score 7; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 LLEPAA 317
    |||||
Db 111 LLEPAA 117

RESULT 17
IF2B_RABIT STANDARD: PRT: 333 AA.
AC P41035;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA).
GN EIF2S2 OR EIF2B.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE;
RX MEDLINE=94032484; PubMed=8218412;
RX Price N.T., Hall L., Proud C.G.;
RT "Cloning of cDNA for the beta-subunit of rabbit translation Initiation

```

```

RT factor-2 using PCR."
RL Biochim. Biophys. Acta 1216:170-172(1993).
RN [2]
RP SEQUENCE OF 25-39, 122-136 AND 195-218.
RX MEDLINE-92002150; PubMed-1911855;
RA Bommer U.A., Kratz R., Kurzhalla T.V., Price N.T., Proud C.G.;
RT "Amino acid sequence analysis of the beta- and gamma-subunits of
  eukaryotic initiation factor eif-2. Identification of regions
  interacting with GTP."
RL Biochim. Biophys. Acta 1079:308-315(1991).
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS
  BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS
  COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING
  TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL
  SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY
  HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP
  BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER
  ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP
  BY WAY OF A REACTION CATALYZED BY EIF-2B.
CC -1- SUBUNIT: HETERODIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
  CHAIN.
CC -1- SIMILARITY: BELONGS TO THE EIF2B / EIF5 FAMILY
  -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
  -----
CC EMBL: X73836; CAA52058.1; -
CC DR PIR: S38536; S38536.
CC DR PIR: S17871; S17871.
CC InterPro: IPR002735; -
CC DR Pfam: PF01873; eIF5_eIF2B; 1.
CC Initiation factor; Protein biosynthesis; Zinc-finger.
CC FT DOMAIN 14 21
CC FT DOMAIN 79 87
CC FT DOMAIN 124 129
CC FT ZN.FING 281 305
CC FT SEQUENCE 333 AA; 38326 MW; 3FFDBAB92DFC1465 CRC64;
SQ

```

Query Match 1.6%; Score 7; DB 1; Length 333;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 291 VOEPAEP 297
   |||||
Db 107 VOEPAEP 113

```

RESULT 18
 CAHC_HUMAN STANDARD; PRT; 354 AA.
 ID CAHC_HUMAN
 AC 043570;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CARBONIC ANHYDRASE XII PRECURSOR (EC 4.2.1.1) (CARBONATE DEHYDRATASE
 XII) (CA-XII) (TUMOR ANTIGEN HOM-RCC-3.1.3).
 GN CA12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 OX [1]
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98301622; PubMed-9636197;
 RA Turci O., Sahin U., Vollmar E., Siemer S., Gottert E., Seitz G.,
 RA Parkkila A.K., Shah G.N., Grubb J.H., Pfundschuh M., Sly W.S.;
 RT "Human carbonic anhydrase XII: cDNA cloning, expression, and

```

RT chromosomal localization of a carbonic anhydrase gene that is
  overexpressed in some renal cell cancers."
RT Proc. Natl. Acad. Sci. U.S.A. 95:7608-7613(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-98445416; PubMed-9770531;
RA Ivanov S.V., Kuzmin I., Wei M.-H., Pack S., Geil L., Johnson B.E.,
  RA Stanbridge E.J., Lerman M.I.;
RT "Down-regulation of transmembrane carbonic anhydrases in renal cell
  carcinoma cell lines by wild-type von Hippel-Lindau transgenes."
RL Proc. Natl. Acad. Sci. U.S.A. 95:12596-12601(1998).
CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN COLON, KIDNEY, AND
  PROSTATE AND MODERATELY EXPRESSED IN PANCREAS, OVARY, AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
  FAMILY.
  -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
  -----
CC EMBL: AF051882; AAC39789.1; -
CC DR EMBL: AF037335; AAC63952.1; -
CC DR MIM: 603263; -
CC InterPro: IPR001148; -
CC DR Pfam: PF00194; carb_anhydrase; 1.
CC DR PROSITE: PS00162; EUR_CO2_ANHYDRASE; 1.
CC KW Lyase; Zinc; Transmembrane; Signal.
CC FT SIGNAL 1 24
CC FT CHAIN 25 354
CC FT DOMAIN 25 301
CC FT TRANSMEM 302 322
CC FT DOMAIN 323 354
CC FT METAL 119 119
CC FT METAL 121 121
CC FT METAL 145 145
CC FT CARBOHYD 28 28
CC FT CARBOHYD 80 80
CC FT CARBOHYD 162 162
CC FT SEQUENCE 354 AA; 39451 MW; 9016216BF2CAGC0C CRC64;
SQ

```

Query Match 1.6%; Score 7; DB 1; Length 354;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 43 AAVLLIV 49
   |||||
Db 9 AAVLLIV 15

```

RESULT 19
 SUR6_HUMAN STANDARD; PRT; 361 AA.
 ID SUR6_HUMAN
 AC 075683; O9UK24;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SURFEIT LOCUS PROTEIN 6.
 GN SURF6 OR SURF-6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 OX [1]
 RN
 RP SEQUENCE FROM N.A.
 RA Magoulas C., Friedl M.;

RT "Isolation and genomic analysis of the human Surf-6 gene: a member of
 the surf6 locus.";
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-31 FROM N.A.
 RA Angelillo A., Russo G., Porcellini A., Smaidone S., Scognamiglio B.,
 RA D'Alessandro F., Pietropaolo C.T.;
 RT "The human homologue of the mouse Surf-5 gene encodes multiple
 RT alternatively spliced transcripts.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN A NUCLEOLAR FUNCTION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEOLAR.
 CC -1- SIMILARITY: BELONGS TO THE SURF6 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF186772; AAC56587.1; -;
 DR EMBL: AJ224639; CAH12054.1; -;
 DR MIM: 183642; -;
 KW Nuclear protein.
 FT DOMAIN 45 49 NUCLEOLAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 154 158 NUCLEOLAR LOCALIZATION SIGNAL (POTENTIAL).
 FT CONFLICT 6 6 A -> T (IN REF. 2).
 SQ SEQUENCE 361 AA; 41449 MW; 5B9639916E882791 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 361;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 PAASGAR 15
 Db 75 PAASGAR 81
 RESULT 20
 METX_CORGL STANDARD; PRT; 379 AA.
 AC 068640;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DE HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31) (HOMOSERINE O-TRANS-
 DE ACETYLASE) (HOMOSERINE TRANSACETYLASE) (HTA).
 GN METX OR META.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=98330894; PubMed=9666465;
 RA Park S.-D., Lee J.-Y., Kim Y., Kim J.-H., Lee H.-S.;
 RT "Isolation and analysis of meta, a methionine biosynthetic gene
 RT encoding homoserine acetyltransferase in corynebacterium glutamicum.";
 RL MCL. Cells 8:286-294(1998).
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + L-HOMOSERINE -> COA + O-ACETYL-L-
 CC HOMOSERINE.
 CC -1- PATHWAY: BIOSYNTHESIS OF METHIONINE; HTA VARIANT; FIRST STEP.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE AB HYDROLASE FAMILY; HTA SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF052652; AAC06035.1; -;
 DR InterPro: IPR000073; -;
 DR Pfam: PF00361; abhydrolase; 1.
 KW Methionine biosynthesis; Transferase; Acyltransferase.
 FT ACT_SITE 145 145
 FT ACT_SITE 343 343 POTENTIAL.
 SQ SEQUENCE 379 AA; 41390 MW; 4AB500D19E649P1 CRC64;
 Query Match 1.6%; Score 7; DB 1; Length 379;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 219 TVAAYVL 225
 Db 135 TVAAYVL 141
 RESULT 21
 CYCH_RHIME STANDARD; PRT; 381 AA.
 ID CYCH_RHIME
 AC P45400;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCH.
 GN CYCH.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AK631;
 RX MEDLINE=95231514; PubMed=7715602;
 RA Kereszt A., Slaska-Kiss K., Putroky P., Banfalvi Z., Kondoros A.;
 RT "The cychK gene of Rhizobium meliloti involved in cytochrome c
 RT biogenesis are regulated for 'respiratory' nitrate reduction ex planta
 RT and for nitrogen fixation during symbiosis.";
 RL MCL. Gen. Genet. 247:39-47(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA Galibert F., Capela D., Hubler-Barloy F., Gatius M., Batut J.,
 RA Bolstead P., Gouzy J., Kahn D., Thebaud P., Goffeau A.,
 RA Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,
 RA Vandenbol M., Puehler A., Becker A., Weidner S.;
 RL Submitted (MAR-2000) to the SWISS-PROT data bank.
 CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
 CC -1- POSSIBLE SUBUNIT OF A HEME LYASE.
 CC -1- SUBCELLULAR LOCATION: MOSTLY PERIPLASMIC; ANCHORED IN THE INNER
 CC MEMBRANE (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X82560; CAA57904.1; -;
 DR Cytochrome c-type biogenesis; Transmembrane; Inner membrane.
 FT TRANSMEM 2 22 POTENTIAL.
 FT TRANSMEM 95 115
 FT DOMAIN 116 381 PERIPLASMIC (POTENTIAL).
 FT CONFLICT 74 74 S -> A (IN REF. 1).
 FT CONFLICT 226 226 R -> P (IN REF. 1).
 FT CONFLICT 240 240 K -> R (IN REF. 1).

```

FT CONFLICT 271 272 EL -> DV (IN REF. 1).
FT CONFLICT 280 281 MISSING (IN REF. 1).
FT CONFLICT 286 286 D -> N (IN REF. 1).
FT CONFLICT 329 329 M -> I (IN REF. 1).
SQ SEQUENCE 381 AA: 40321 MW: DFFB39391554033E CRC64;

Query Match 1.6%; Score 7; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 VAAVLL 48
DB 14 VAAVLL 20

RESULT 22
NR23_MOUSE STANDARD; PRT; 395 AA.
ID NR23_MOUSE
AC 09QXZ7;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PHOTORECEPTOR-SPECIFIC NUCLEAR RECEPTOR (RETINA-SPECIFIC NUCLEAR
DE RECEPTOR).
GN NR23 OR PNR OR RMR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Retina;
MEDLINE=20079622; PubMed=10611353;
RA Chen F., Figueroa D.J., Marmorstein A.D., Zhang Q., Petrukhin K.,
RA Caskey C.T., Austin C.P.;
RA "Retina-specific nuclear receptor: A potential regulator of cellular
RA retinaldehyde-binding protein expressed in retinal pigment epithelium
RA and Muller glial cells."
RL Proc. Natl. Acad. Sci. U.S.A. 96:15149-15154(1999).
[2]
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6J; TISSUE-Retina;
MEDLINE=20266403; PubMed=10805811;
RA Akhmedov N.B., Piriev N.I., Chang B., Rapoport A.L., Hawes N.L.,
RA Nishina P.M., Nustimowitz S., Heckenlively J.R., Roderick T.H.,
RA Kozak C.A., Danciger M., Davison M.T., Farber D.B.;
RA "A deletion in a photoreceptor-specific nuclear receptor mRNA causes
RA retinal degeneration in the rd7 mouse."
RL Proc. Natl. Acad. Sci. U.S.A. 97:5551-5556(2000).
RT - FUNCTION: TRANSCRIPTION FACTOR THAT MAY HAVE A ROLE IN THE
RT REGULATION OF SIGNALING PATHWAYS INTRINSIC TO THE PHOTORECEPTOR
RT CELL FUNCTION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: RETINA-SPECIFIC.
CC -1- DISEASE: DEFECTS IN THIS PROTEIN ARE THE CAUSE OF RETINAL
CC DEGENERATION.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF148129; AAF22228.1; -
CC EMBL: AF204053; AAF69682.1; -
CC MGD: MGI:1346317; Nr2e3.
CC InterPro: IPR000003; -
CC InterPro: IPR000324; -

```

```

DR InterPro: IPR000536; -
DR InterPro: IPR001628; -
DR Pfam: PF00104; hormone_rec. 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00047; STROIDINGER.
DR PRINTS: PR00350; VITAMINDR.
DR PRINTS: PR00545; RETINOIDXR.
DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
FT ZINC_BIND 40 107 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 40 60 C4-TYPE.
FT ZN_FING 76 102 C4-TYPE.
SQ SEQUENCE 395 AA: 43176 MW: D8E391CE24919A68 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 205 PASPCL 211
DB 198 PASPCL 204

RESULT 23
LEF1_MOUSE STANDARD; PRT; 397 AA.
ID LEF1_MOUSE
AC P27782;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LYMPHOID ENHANCER BINDING FACTOR 1 (LEF-1).
GN LEF1 OR LEF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA;
MEDLINE=91224496; PubMed=1827423;
RA Trivis A., Amsterdam A., Belanger C., Grosschedl R.;
RA "LEF-1, a gene encoding a lymphoid-specific protein with an HMG
RA domain, regulates T-cell receptor alpha enhancer function."
RL Genes Dev. 5:880-894(1991).
[2]
RN STRUCTURE BY NMR OF 296-380.
RX MEDLINE=95379976; PubMed=7651541;
RA Love J.J., Li X., Case D.A., Gliese K., Grosschedl P., Wright P.E.;
RT "Structural basis for DNA bending by the architectural transcription
RT factor LEF-1."
RL Nature 376:791-795(1995).
[3]
RN STRUCTURE BY NMR OF 296-380.
RA Li X., Love J.J., Case D.A., Wright P.E.;
RT "High resolution NMR structure of the LEF-1 HMG domain complexed with
RT its cognate DNA."
RL Submitted (OCT-1998) to the PDB data bank.
CC -1- FUNCTION: REGULATES T-CELL RECEPTOR ALPHA ENHANCER FUNCTION.
CC BINDS DNA IN A SEQUENCE-SPECIFIC MANNER.
CC -1- SUBUNIT: INTERACTS WITH BETA-CATENIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: LYMPHOCYTES.
CC -1- DOMAIN: PROLINE-RICH AND ACIDIC REGIONS ARE IMPLICATED IN THE
CC ACTIVATION FUNCTIONS OF RNA POLYMERASE II TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: X58636; CAA1493.1; -
CC PIR: A39565; A39565.
DR PIR; 2LEF; 21-OCT-98.
DR TRANSFAC; T00930; -
DR MGD; MGI:96770; LeF1.
DR InterPro: IPR000910; -
DR Pfam: PF00505; HMG_box; 1.
KM DNA-binding; Nuclear protein; Transcription regulation; 3D-structure.
FT DOMAIN 6 11
FT POLY-GLY.
FT DOMA1N 12 50
FT ASP/GLU-RICH (ACIDIC).
FT DOMA1N 75 271
FT PRO-RICH.
FT DNA BIND 297 365
FT HMG BOX.
FT DOMA1N 372 377
FT POLY-LYS.
SQ SEQUENCE 397 AA; 44058 MW; 0CA01B6C528FD7FA CRC64;

Query Match 1.6%; Score 7; DB 1; Length 397;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 GGGGDPE 256
|||||
DB 8 GGGGDPE 14

RESULT 24

LEF1_RAT STANDARD; PRT; 397 AA.

AC 090XN1;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LYMPHOID ENHANCER BINDING FACTOR 1 (LEF-1).
GN LEF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Kidney;
RA Kobiela K., Kobiela A., Trzeciak W.H.;
RT "Cloning of lymphoid enhancer binding factor-1 (LEF-1) from rat kidney;
RT homology to the murine sequence.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REGULATES T-CELL RECEPTOR ALPHA ENHANCER FUNCTION.
CC BINDS DNA IN A SEQUENCE-SPECIFIC MANNER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AF198533; AAF15601.1; -
KM DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 6 11
FT POLY-GLY.
FT DOMA1N 12 50
FT ASP/GLU-RICH (ACIDIC).
FT DOMA1N 75 271
FT PRO-RICH.
FT DNA BIND 297 365
FT HMG BOX.
FT DOMA1N 372 377
FT POLY-LYS.
SQ SEQUENCE 397 AA; 44023 MW; D6BEFF805CB526EA CRC64;

Query Match 1.6%; Score 7; DB 1; Length 397;
Best Local Similarity 100.0%; Pred. No. 54;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 GGGGDPE 256
|||||
DB 8 GGGGDPE 14

RESULT 25

LEF1_HUMAN STANDARD; PRT; 399 AA.

AC 090U2;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LYMPHOID ENHANCER BINDING FACTOR 1 (LEF-1) (T CELL-SPECIFIC
DE TRANSCRIPTION FACTOR 1-ALPHA) (TCF1-ALPHA).
GN LEF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kobiela K., Kobiela A., Trzeciak W.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REGULATES T-CELL RECEPTOR ALPHA ENHANCER FUNCTION.
CC BINDS DNA IN A SEQUENCE-SPECIFIC MANNER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AF198532; AAF13268.1; -
DR DR MTM; 153245; -
KM DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 6 13
FT POLY-GLY.
FT DOMA1N 14 52
FT ASP/GLU-RICH (ACIDIC).
FT DOMA1N 77 273
FT PRO-RICH.
FT DNA BIND 299 367
FT HMG BOX.
FT DOMA1N 374 379
FT POLY-LYS.
SQ SEQUENCE 399 AA; 44200 MW; D480D440698EEF3 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 GGGGDPE 256
|||||
DB 10 GGGGDPE 16

RESULT 26

Y413_ARATH STANDARD; PRT; 435 AA.

AC 004656;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 48.8 KDA PROTEIN M021B04.13.
GN A_TMO21B04.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RC Dante M., Wamsley P., Gibson A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE NOPS/NOP56 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF007271; AAB61074.1; -
DR InterPro: IPR002687; -
DR Pfam: PF01798; NOP: 1.
KW Hypothetical protein.
SQ SEQUENCE 435 AA; 48818 MW; BE502C8C244AF085 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 435;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 174 SDIECVH 180
| | | | |
Db 45 SDIECVH 51

RESULT 27

SYD_METH
ID SYD_METH STANDARD; PRT; 437 AA.
AC 026328;

DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 15-JUL-1998 (rel. 36, Last annotation update)
DE ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12) (ASPARTATE--TRNA LIGASE)
DE (ASPRS).
DE GN
GN ASPs OR MTH226.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacter.
OC NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Safer H., Patweli D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delat: functional analysis and comparative genomics.";
RT J. Bacteriol. 179:7135-7155(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +
CC PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000809; AAB84732.1; -
DR HSSP: P04802; IAST.

DR InterPro: IPR002106; -
DR InterPro: IPR002309; -
DR InterPro: IPR002312; -
DR Pfam: PF00152; tRNA-synt.2; 1.
DR PRINTS: PR01042; TRNASYNTHASP.
DR PROSITE: PS00179; AA-TRNA_LIGASE_II_1; FALSE_NEG.
DR PROSITE: PS00359; AA-TRNA_LIGASE_II_2; FALSE_NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ SEQUENCE 437 AA; 49838 MW; 128C9BE08FBFA6D3 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 437;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 404 DALETIG 410
| | | | |
Db 268 DALETIG 274

RESULT 28

RECN_HELPJ
ID RECN_HELPJ STANDARD; PRT; 522 AA.
AC 092J80;

DT 01-OCT-2000 (rel. 40, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N).
DE RECN OR JHP1434.
GN Helicobacter pylori J99 (Campylobacter pylori J99).
OS Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OC NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.

RA MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moll D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guille B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Mickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RT Nature 397:176-180(1999).
CC -1- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
CC DNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RECN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001566; AAD07019.1; -
DR InterPro: IPR001617; -
DR NP_BIND: ATP-binding.
KW NP_BIND 36
SQ SEQUENCE 522 AA; 59349 MW; 000D7EB3C0632EAS CRC64;

Query Match 1.6%; Score 7; DB 1; Length 522;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 402 LDDALET 408
| | | | |
Db 156 LDDALET 162

RESULT 29

EAT3_MOUSE
ID EAT3_MOUSE STANDARD: PRT: 523 AA.
AC P51906; 035869; 035875;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENT
GLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY AMINO-ACID CARRIER 1).
GN SLC1A1 OR EAT3 OR EAAC1.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCL:ICR; TISSUE=Cerebellum;
MEDLINE=95284091; PubMed=7766664;
RA Mukaiwa Y., Tanaka K., Hagiwara T., Wada K.;
RT "Molecular cloning of two glutamate transporter subtypes from mouse
brain.";
RL Blochm. Biophys. Acta 1244:233-237(1995).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
PENG J.-B., Guo L.-H.;
Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG, KIDNEY AND SKELETAL
MUSCLE.
CC -1- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTER FAMILY
(SDF).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: D43797; BAA07855.1; -
DR EMBL: U73521; AAB0694.1; -
DR EMBL: U75217; AAB71742.1; -
DR MGD; MGI:105083; Slc1a1.
DR InterPro: IPR001991; -
DR Pfam: PF00375; SDF: 1.
DR PRINTS: PR00173; EDTNSPORT.
DR PROSITE: PS00713; NA_DICARBOXYL_SYM_1; 1.
DR PROSITE: PS00714; NA_DICARBOXYL_SYM_2; 1.
KW Transport; Transmembrane; Glycoprotein; Symport; multigene family.
FT DOMAIN 1 18
FT TRANSMEM 19 38
FT TRANSMEM 62 82
FT TRANSMEM 94 114
FT TRANSMEM 115 208
FT DOMAIN 209 228
FT TRANSMEM 253 272
FT TRANSMEM 289 308
FT TRANSMEM 315 334
FT TRANSMEM 361 380
FT TRANSMEM 391 410
FT TRANSMEM 419 438
FT TRANSMEM 438 458
FT CARBOHYD 128 138
FT CARBOHYD 178 178
FT CARBOHYD 194 194
FT CARBOHYD 121 121
FT CONFLICT 121 121
FT CONFLICT 139 139
FT CONFLICT I -> A (IN AAB71742).
FT CONFLICT I -> M (IN REF. 2).
SQ SEQUENCE 523 AA; 56675 MW; F20AD35447F398FB CRC64;

Query Match 1.6%; Score 7; DB 1; Length 523;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 219 TVAAVVL 225
Db 24 TVAAVVL 30
RESULT 30
EAT3_MOUSE
ID EAT3_MOUSE STANDARD: PRT: 523 AA.
AC P51907; 063727;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENT
GLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY AMINO-ACID CARRIER 1).
GN SLC1A1 OR EAT3 OR EAAC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Bjors M., Gjesdal O., Erickson J.D., Torp R., Levy L.M.,
RA Ohtsen O.P., Degre M., Storm-Mathisen J., Seeberg E.,
RA Danbolt N.C.;
Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
MEDLINE=96318324; PubMed=8747133;
RA Kanai Y., Bhile P.G., Dilligila M., Hediger M.A.;
RT "Neuronal high-affinity glutamate transport in the rat central
nervous system.";
RL NeuroReport 6:2357-2362(1995).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RA Kiyu S., Yao G.L., Morita N., Kato H., Kiyama H.;
Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 156-368 FROM N.A.
MEDLINE=96346642; PubMed=8738164;
RA Palos T.P., Ramachandran B., Boado R., Howard B.D.;
RT "Rat C6 and human astrocytic tumor cells express a neuronal type of
glutamate transporter.";
RL Brain Res. Mol. Brain Res. 37:297-303(1996).
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTER FAMILY
(SDF).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: X94255; CAA63937.1; -
DR EMBL: U39555; AAB09773.1; -
DR EMBL: D63772; BAA09849.1; -
DR EMBL: L35358; AAB51161.1; -
DR InterPro: IPR001991; -
DR Pfam: PF00375; SDF: 1.

DR PRINTS: PRO0173; EDTNSPORT.
 DR PROSITE: PS00713; NA.DICARBOXYL_SYM_1; 1.
 DR TRANSPORT; Transmembrane; Glycoprotein; Symport; Multigene family.
 KW DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 19 38 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 115 208 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 209 228 POTENTIAL.
 FT TRANSMEM 253 272 POTENTIAL.
 FT TRANSMEM 289 308 POTENTIAL.
 FT TRANSMEM 315 334 POTENTIAL.
 FT TRANSMEM 361 380 POTENTIAL.
 FT TRANSMEM 391 410 POTENTIAL.
 FT TRANSMEM 419 438 POTENTIAL.
 FT CARBOHYD 128 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 68 68 L -> M (IN REF. 3).
 FT CONFLICT 93 93 R -> C (IN REF. 3).
 FT CONFLICT 149 151 MFP -> ILG (IN REF. 2).
 FT CONFLICT 156 156 Q -> E (IN REF. 4).
 FT CONFLICT 203 203 G -> C (IN REF. 4).
 FT CONFLICT 218 218 C -> F (IN REF. 4).
 FT CONFLICT 309 309 R -> T (IN REF. 4).
 FT CONFLICT 383 383 L -> V (IN REF. 3).
 SQ SEQUENCE 523 AA; 56771 MW; D6783244E37640E4 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 523;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 219 TVAAVYL 225
 |||||
 DB 24 TVAAVYL 30

RESULT 31
 EAT3_HUMAN STANDARD; PRT; 524 AA.
 AC P43005;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENT
 GN GLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY AMINO-ACID CARRIER 1).
 DE SLC1A1 OR EAT3 OR EAMC1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain cortex;
 RX MEDLINE=94365697; PubMed=7521911;
 RA Artiza J.L., Fairman W.A., Wendy A., Wadiche J.I., Murdoch G.H.,
 RA Kavanaugh M.P., Amara S.G.;
 RT "Functional comparisons of three glutamate transporter subtypes
 RT cloned from human motor cortex."
 RL J. Neurosci. 14:5559-5569(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95162720; PubMed=7859077;
 RA Shashidharan P., Huntley G.W., Meyer T., Morrison J.H.,
 RA Platakis A.;
 RT "Neuron-specific human glutamate transporter: molecular cloning,
 RT characterization and expression in human brain."
 RL Brain Res. 662:245-250(1994).
 RN [3]
 RP SEQUENCE FROM N.A.

RX MEDLINE=94327639; PubMed=7914198;
 RA Kanai Y., Steizner M., Nussberger S., Khawaja S., Hebert S.C.,
 RA Smith C.P., Hediger M.A.;
 RT "The neuronal and epithelial human high affinity glutamate
 RT transporter. Insights into structure and mechanism of transport."
 J. Biol. Chem. 269:20599-20606(1994).
 CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
 CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
 CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
 CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- PTM: GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTER FAMILY
 CC (SDF).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U03506; AAA50430.1; -;
 CC EMBL: U06469; AAA53215.1; -;
 CC MIM: 133550; -;
 DR InterPro: IP001991; -;
 DR Pfam: PF00375; SDF; 1.
 DR PRINTS: PRO0173; EDTNSPORT.
 DR PROSITE: PS00713; NA.DICARBOXYL_SYM_2; 1.
 DR TRANSPORT; Transmembrane; Glycoprotein; Symport; Multigene family.
 KW DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 19 38 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 115 209 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 210 229 POTENTIAL.
 FT TRANSMEM 254 273 POTENTIAL.
 FT TRANSMEM 290 309 POTENTIAL.
 FT TRANSMEM 316 335 POTENTIAL.
 FT TRANSMEM 362 381 POTENTIAL.
 FT TRANSMEM 392 411 POTENTIAL.
 FT TRANSMEM 420 439 POTENTIAL.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 10 10 E -> PS (IN REF. 1).
 FT CONFLICT 94 94 A -> V (IN REF. 2).
 FT CONFLICT 96 96 V -> L (IN REF. 3).
 FT CONFLICT 170 170 K -> N (IN REF. 3).
 FT CONFLICT 172 172 P -> A (IN REF. 3).
 FT CONFLICT 260 260 L -> H (IN REF. 3).
 FT CONFLICT 280 280 R -> A (IN REF. 2).
 SQ SEQUENCE 524 AA; 57100 MW; FC6244ADC9EA228F CRC64;

Query Match 1.6%; Score 7; DB 1; Length 524;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 219 TVAAVYL 225
 |||||
 DB 24 TVAAVYL 30

RESULT 32
 CH61_MYCTU STANDARD; PRT; 539 AA.
 ID CH61_MYCTU
 AC Q59573; Q59581;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 60 KDA CHAPERONIN 1 (PROTEIN CPN60-1) (GROEL PROTEIN 1).
GN GROEL1 OR GROEL-1 OR RV3417C OR MTCV78.12.
OS Mycobacterium tuberculosis.
OC Bacteria: Firmicutes; Actinobacteriae; Actinobacteridae;
OC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jajuel K., Krogan A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=ERDMANN;
RX MEDLINE=93219332; PubMed=7681982;
RA Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;
RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs."
RL Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612(1993).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC -1- 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z77165; CAB01006.1; -
DR EMBL: X60350; CAA42909.1; -
DR HSSP: P06139; IGRU.
DR TubercuList: RV3417C; -
DR InterPro: IPR001844; -
DR Pfam: PF00118; cpn60_TCP1.1
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60.1.
DR KMAP: ATP-binding; Multigene family: Antigen: Cell wall.
FT VARIANT 467 467 N -> K (IN STRAIN ERDMANN).
FT ACT SITE 521 521
SQ SEQUENCE 539 AA: 55877 MW: 3E0B93164C091B63 CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 539;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 409 LGERLAK 415
Db 363 LGERLAK 369

```

```

RESULT 33
GSHC_TOBAC STANDARD; PRT; 557 AA.
ID GSHC_TOBAC
AC P80461;
DT 01-OCT-1996 (Rel. 34, Created)

```

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTATHIONE REDUCTASE, CHLOROPLAST PRECURSOR (EC 1.6.4.2) (GR)
DE (GRASE) (FRAGMENT).
GN Nicotiana tabacum (Common tobacco).
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons: core eudicots: Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=96081527; PubMed=8547821;
RA Creissen G.P., Mullineaux P.M.;
RT "Cloning and characterisation of glutathione reductase cDNAs and
RT identification of two genes encoding the tobacco enzyme."
RL Planta 197:422-425(1995).
RN 12
RP SEQUENCE OF 67-74.
RC TISSUE=Leaf;
RA Willows R.D., Kannangara G.C., Svendsen I.;
RT Submitted (JUN-1995) to the SWISS-PROT data bank.
CC -1- FUNCTION: MAINTAIN HIGH LEVELS OF REDUCED GLUTATHIONE IN THE
CC CHLOROPLAST.
CC -1- CATALYTIC ACTIVITY: NADPH + OXIDIZED GLUTATHIONE = NADP(+) +
CC 2 GLUTATHIONE.
CC -1- COFACTOR: FAD.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X76293; CAA53925.1; -
DR HSSP: P00390; IALG.
DR InterPro: IPR001100; -
DR Pfam: PF00070; PYR_redox.1.
DR PROSITE: PS00076; PYRIDINE_REDOX.1.
DR Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP;
KW Chloroplast; Transil peptide.
FT TRANSIT 1 1
FT NON_TER 1
FT TRANSIT <1 66 CHLOROPLAST.
FT CHAIN 67 557 GLUTATHIONE REDUCTASE.
FT NP_BIND 81 111 FAD (ADP PART) (BY SIMILARITY).
FT DISULFID 127 132 REDOX-ACTIVE (BY SIMILARITY).
FT BINDING 283 283 NADP (BY SIMILARITY).
FT BINDING 289 289 NADP (BY SIMILARITY).
FT NP_BIND 378 388 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT SITE 521 521
SQ SEQUENCE 557 AA: 60035 MW: 79813D1C6A1D784 CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 557;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 73 RSPSEG 79
Db 539 RSPSEG 545

```

```

RESULT 34
ROL_HUMAN STANDARD; PRT; 558 AA.
ID ROL_HUMAN
AC P14866;
DT 01-APR-1990 (Rel. 14, Created)

```

DT 01-APR-1990 (Rel. 14, last sequence update)
 DT 01-OCT-2000 (Rel. 40, last annotation update)
 DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L (HNRL)
 GN HNRL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90078296; PubMed=2687284;
 RA Pincil-Roma S., Swanson M.S., Gall J.G., Dreyfuss G.;
 RT "A novel heterogeneous nuclear RNP protein with a unique distribution
 on nascent transcripts";
 RL J. Cell Biol. 109:2575-2587(1989).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=93162043; PubMed=1286667;
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Cells J.E.,
 RA Vandekerckhove J.;
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel
 protein database of normal human epidermal keratinocytes.";
 RL Electrophoresis 13:960-969(1992).
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE HETEROGENEOUS NUCLEAR
 RIBONUCLEOPROTEIN (HNRP) COMPLEXES WHICH PROVIDE THE SUBSTRATE
 FOR THE PROCESSING EVENTS THAT PRE-MRNAs UNDERGO BEFORE BECOMING
 FUNCTIONAL, TRANSLATABLE MRNAs IN THE CYTOPLASM. L IS ASSOCIATED
 WITH MOST NASCENT TRANSCRIPTS INCLUDING THOSE OF THE LANDMARK
 CC GIANT LOOPS OF AMPHIBIAN LAMPBRUSH CHROMOSOMES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOPLASM.
 CC -1- PTR: SEVERAL ISOELECTRIC FORMS OF THE L PROTEIN ARE PROBABLY THE
 RESULTS OF POSTTRANSLATIONAL MODIFICATIONS.
 CC -1- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM)
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16135; CAA34261.1; -
 DR PIR: A33616; A33616.
 DR HSSP: P19339; 2SXL.
 DR SMSS-2DPAGE; P14866; HUMAN.
 DR Aairus/Ghent-2DPAGE; 1305; IEF.
 DR Aairus/Ghent-2DPAGE; 4602; NEPHGE.
 DR MIM: 603083; -
 DR MIM: 164021; -
 DR Interpro: IPR000504; -
 DR Pfam: PF00076; rrm; 3.
 DR PROSITE: PS50102; RRM; 3.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE NEG.
 DR Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
 FT DOMAIN 8 58
 FT DOMAIN 71 145 RNA-BINDING (RRM) 1.
 FT DOMAIN 162 239 RNA-BINDING (RRM) 2.
 FT DOMAIN 351 425 RNA-BINDING (RRM) 3.
 FT DOMAIN 335 342 POLY-PRO.
 SQ SEQUENCE 558 AA; 60187 MW; 395E5A04B14C848D CRC64;

Query Match 1.6%; Score 7; DB 1; Length 558;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 DALETIG 410
 DB 523 DALETIG 529

RESULT 35
 CCMF_PSEFL
 ID CCMF_PSEFL STANDARD; PRT; 660 AA.
 AC P52225;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYC.
 GN CYC.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OC NCBI_Taxid=294;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=09906;
 RA Yang C.H., Azad H.R., Cooksey D.A.;
 RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMES.
 CC POSSIBLE SUBUNIT OF A HEME LYASE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CCMF/CYC/CCL/NRFE/CSEA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U44827; AAC44226.1; -
 DR Interpro: IPR002541; -
 DR Pfam: PF01578; CytoC_asm; 1.
 DR Cytochrome c-type biogenesis; Transmembrane; Inner membrane.
 KW Transmem 15 35
 FT TRANSMEM 49 119 POTENTIAL.
 FT TRANSMEM 99 169 POTENTIAL.
 FT TRANSMEM 126 146 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 216 236 POTENTIAL.
 FT TRANSMEM 250 270 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 360 380 POTENTIAL.
 FT TRANSMEM 395 415 POTENTIAL.
 FT TRANSMEM 426 446 POTENTIAL.
 FT TRANSMEM 450 470 POTENTIAL.
 FT TRANSMEM 493 513 POTENTIAL.
 FT TRANSMEM 618 638 POTENTIAL.
 SQ SEQUENCE 660 AA; 72172 MW; 3C6F4B578C56436B CRC64;

Query Match 1.6%; Score 7; DB 1; Length 660;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 LVLVVA 44
 DB 361 LVLVVA 367

RESULT 36
 P115_CHICK
 ID P115_CHICK STANDARD; PRT; 762 AA.
 AC Q98917;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MELANOSOMAL MATRIX PROTEIN 115 KDA PROTEIN PRECURSOR.
 GN MPM115.
 OS Gallus gallus (Chicken).

CC	Euryarchaeota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;
OC	Aurogasteria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBL_TaxID=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-WHITE LEGRON; TISSUE-retinal pigment epithelium;
RX	MELLINE-92020667; PubMed-1924173;
RA	Mochil M., Agata K., Eguchi G.;
RT	"Complete sequence and expression of a cDNA encoding a chicken
RL	115-kDa melanosomal matrix protein."
RL	Pigment Cell Res. 4:41-47(1991).
RP	[2]
RP	CHARACTERIZATION.
RC	STRAIN-WHITE LEGRON; TISSUE-retinal pigment epithelium;
RY	MELLINE-88311098; PubMed-3409326;
RA	Mochil M., Agata K., Kobayashi H., Yamamoto T.S., Eguchi G.;
RT	"Expression of gene coding for a melanosomal matrix protein
RT	transcriptionally regulated in the transdifferentiation of chick
RT	embryo pigmented epithelial cells.";
RL	Cell Differ. 24:67-74(1988).
CC	-1- FUNCTION: MIGHT BE REQUIRED FOR POLYMERIZATION OF MELANIN ONTO THE
CC	CORE STRUCTURE OF MELANOSOMES WITH ENZYMIC FUNCTION OF TYROSINASE.
CC	-1- SUBCELLULAR LOCATION: ON THE FIBROUS MATRIX STRUCTURE OF THE
CC	PREMELANOSOME.
CC	-1- TISSUE SPECIFICITY: SPECIFIC TO PIGMENTED EPITHELIAL CELLS AND
CC	MELANOCYTES. NOT EXPRESSED IN LENS, NEURAL RETINA, BRAIN, HEART,
CC	GIZZARD OR LIVER.
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE REDIFFERENTIATION OF
CC	PIGMENTED EPITHELIAL CELLS (PEC).
CC	-1- PTM: GLYCOSYLATED.
CC	-1- SIMILARITY: BELONGS TO THE PMEL-17/MMB FAMILY.
CC	-1- SIMILARITY: CONTAINS 1 PKD DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; D88348; BAA13589.1; .
DR	InterPro; IPR000601; .
DR	Pfam; PF00801; PKD. 1.
DR	PROSITE; PSS0093; PKD. 1.
KW	Signal; Glycoprotein; Repeat.
FT	SIGNAL
FT	1 19
FT	CHAIN 20 762
FT	19
FT	MELANOSOMAL MATRIX PROTEIN 115 KDA
FT	PROTEIN.
FT	DOMAIN 223 323
FT	REPEAT 441 532
FT	REPEAT 441 464
FT	REPEAT 465 488
FT	REPEAT 489 508
FT	REPEAT 509 532
FT	REPEAT 511 532
FT	CARBOHYD 111 111
FT	CARBOHYD 115 115
FT	CARBOHYD 346 346
FT	CARBOHYD 651 651
FT	CARBOHYD 651 651
FT	CARBOHYD 659 659
FT	CARBOHYD 752 AA; 77356 MW; 172C8DBDAEFDCE76G CXC64;
SO	SEQUENCE
FT	PKD.
FT	4 X 20-24 AA APPROXIMATE TANDEM REPEATS.

Query Match	1.6%	Score 7:	DB 1:	Length 762:
Best Local Similarity	100.0%	Pred. No. 94:		
Matches	7:	Conservative	0:	Mismatches
			0:	Indels
			0:	Gaps
OY	294	PAEPTGV	300	
db	368	PAEPTGV	374	

RESULT	37	SC18_CANAL	STANDARD:	PRF:	794 AA.
ID	SC18_CANAL				
AC	P34732;				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	VESICULAR-FUSION PROTEIN SEC18.				
GN	SEC18.				
OS	Candida albicans (Yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
CC	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
OX	NCBI_TaxID=5476;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 26355;				
RX	MEDLINE=94025979; PubMed=8212895;				
RA	Nieto A., Sanz P., Sentandreu R., del Castillo Agudo L.;				
RT	"Cloning and characterization of the SEC18 gene from Candida albicans ";				
RL	Yeast 9:875-887(1993).				
CC	-I- FUNCTION: REQUIRED FOR VESICLE-MEDITED TRANSPORT. CATALYZES THE				
CC	POSITION OF TRANSPORT VESICLES WITHIN THE GOLGI CISTERNAE. IS ALSO				
CC	REQUIRED FOR TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO THE GOLGI				
CC	STACK. SEEM TO FUNCTION AS A FUSION PROTEIN REQUIRED FOR THE				
CC	DELIVERY OF CARBO PROTEINS TO ALL COMPARTMENTS OF THE GOLGI STACK				
CC	INDEPENDENT OF VESICLE ORIGIN. BINDS TO SEC17.				
CC	-I- SUBCELLULAR LOCATION: CYTOPLASMIC.				
CC	-I- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: X66467; CAA47077.1; -.				
DR	PIR: S37606; S37606.				
DR	InterPro: IPR001939; -.				
DR	Pfam: PF00004; AAA; 1.				
DR	PROSITE: PS00674; AAA; 1.				
KW	Transport; Protein transport; Endoplasmic reticulum; Golgi stack;				
KW	ATP-binding; Repeat.				
FT	NP_BIND 317 324				
FT	ATP (POTENTIAL).				
FT	NP_BIND 600 607				
FT	ATP (POTENTIAL).				
SO	SEQUENCE 794 AA; 88925 MW; 35A22B62C196F67D CRC64;				

```

Query Match Similarity 1.6%; Score 7; DB 1; Length 794;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 278 ILQPROV 284
      |||||
      |||||
      |||||

Db 173 ILQPROV 179

RESULT 38
RRPO_SBMV STANDARD: PRT: 956 AA.
ID RRPO_SBMV STANDARD: PRT: 956 AA.
AC P21403;
DC 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PROBABLE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48).
DS Southern bean mosaic virus (SBMV).
OS Viruses: SSRNA positive-strand viruses, no DNA stage; Sobemovirus.
OC NCBI_TaxID=12139;
OX NCBI_TaxID=12139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COMPEA:

```



```

RX MEDLINE=88044510; PubMed=2823471;
RA Mu S., Rinehart C.A., Kaesberg P.;
RT "Sequence and organization of southern bean mosaic virus genomic
RL RNA."
RL Virology 161:73-80(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M23021; AAA46565.1; -
DR PIR; B37379; RREMSC.
DR InterPro; IPR001795; -
DR Pfam; PF02123; Luteo_ORF3; 1.
DR PRINTS; PR00914; LVIRUSRNAPOL.
KM Transferase; RNA-directed RNA polymerase.
SO SEQUENCE 956 AA; 104844 MW; ABB981C9B76082D8 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 956;
Best Local Similarity 100.0%; Pred.No. 1,le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 199 TSSPCTP 205
DB 516 TSSPCTP 522
IIIIIIII

RESULT 39
RBP2_DROME STANDARD; PRT; 1176 AA.
ID RBP2_DROME STANDARD; PRT; 1176 AA.
AC P08266; G95027; Q04155; Q9VFW7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II 140 KDA POLYPEPTIDE (EC 2.7.7.6)
DE (RNA POLYMERASE II SUBUNIT 2).
GN RPII140 OR CG3180.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE OF 54-1176 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=88011299; PubMed=3116266;
RA Falkenburg D., Dornick B., Faust D.M., Bautz E.K.F.;
RT "RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit
RT to the beta subunit of Escherichia coli RNA polymerase."
RJ J. Mol. Biol. 195:929-937(1987).
RN [2]
RP SEQUENCE OF 1-69 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=91276237; PubMed=1905256;
RA Stizler S., Oldenburg I., Peterson G., Bautz E.K.F.;
RT "Analysis of the promoter region of the housekeeping gene DmRPI40 by
RT sequence comparison of Drosophila melanogaster and Drosophila
RT virilis."
RL Gene 100:155-162(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek J.R., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrita J.F., Abayant A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brooksstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matell B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splitter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -I- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 14
CC DIFFERENT POLYPEPTIDES. RNA POLYMERASE II CONSISTS OF 10 DIFFERENT
CC SUBUNITS. THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA
CC POLYMERASE II.
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -I- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X05709; CAA29180.2; -
DR EMBL; M62972; AAA28476.1; -
DR EMBL; AE003703; AAF55024.1; -
DR PIR; A27826; A27826.
DR FLYBase; FBgn0003276; RPII140.
DR InterPro; IPR001572; -
DR Pfam; PF00562; RNA_POL_B; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KM Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
RN Zinc-finger; Metal-binding; Nuclear protein.
FT ZN.FING 1121 1142 C4-TYPE (POTENTIAL).
FT CONFLICT 72 72 A -> R (IN REF. 1).
FT CONFLICT 666 667 ID -> MY (IN REF. 1).
SO SEQUENCE 1176 AA; 134042 MW; 224821B335BED7F0 CRC64;

```


Query Match
Best Local Similarity 1.6%; Score 7; DB 1; Length 1176;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 ARGAPG 31
|||||||

DB 324 ARGAPG 330

RESULT 40
POL_SINDO STANDARD; PRT; 1245 AA.

AC P27285; 000349;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE STRUCTURAL POLYPEPTIDE (P130) (CONTAINS: COAT PROTEIN C (EC 3.4.21.-));
DE SPIKE GLYCOPROTEIN E2, E2 AND E1; 6 KDA PEPTIDE;
OS Sindbis virus (subtype Ockelbo / strain Edsbyn 82-5);
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11699;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91220725; PubMed-1673813;
RA Shitko Y., Nilsson B., Dalrymple J.M., Strauss E.G., Strauss J.H.;
RT "Structure of the Ockelbo virus genome and its relationship to other
RT Sindbis viruses.";
RT Virology 182:753-764(1991).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE SPIKE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HEMAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M69205; AAA36973.1; -
CC EMBL; M69207; AAA73066.1; -
CC PIR; B39991; VHWB82.
DR HSP; P03316; IKXF.
DR MEROPS; S03.001; -
DR InterPro; IPR000930; -
DR InterPro; IPR000936; -
DR InterPro; IPR001836; -
DR InterPro; IPR002533; -
DR InterPro; IPR002548; -
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR Pfam; PF00944; Alpha_core; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Coat protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
KW Serine protease.
FT CHAIN 1 264 COAT PROTEIN C (CAPSID PROTEIN C).
FT CHAIN 265 328 SPIKE GLYCOPROTEIN E3.
FT CHAIN 329 751 SPIKE GLYCOPROTEIN E2.
FT CHAIN 752 806
FT CHAIN 807 1245 SPIKE GLYCOPROTEIN E1.
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 147 147 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 223 223 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT TRANSMEM 696 712 POTENTIAL.
FT TRANSMEM 728 746 POTENTIAL.
FT TRANSMEM 768 784 POTENTIAL.
FT TRANSMEM 786 802 POTENTIAL.

FT TRANSMEM 1216 1234 POTENTIAL.
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1051 1051 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1245 AA; 136649 MW; 967EF00E675F84EF CRC64;

Query Match
Best Local Similarity 1.6%; Score 7; DB 1; Length 1245;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 216 IGVTVAA 222
|||||||

DB 708 IGVTVAA 714

RESULT 41
BIMB_EMEI
ID BIMB_EMEI STANDARD; PRT; 2067 AA.

AC P33144;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE CELL DIVISION-ASSOCIATED PROTEIN BIMB.
GN BIMB.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiaceae; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=A773.
RX MEDLINE-92348436; PubMed-1639810;
RA "May G.S., McGoldrick C.A., Holt C.L., Denison S.H.;
RT "The bimB mutation of Aspergillus nidulans uncouples DNA replication
RT from the completion of mitosis.";
RL J. Biol. Chem. 267:15737-15743(1992).
CC -1- FUNCTION: REQUIRED FOR NUCLEAR DIVISION. COULD FUNCTION IN THE
CC MITOTIC SPINDLE.
CC
CC SUBCELLULAR LOCATION: NUCLEAR.
CC
CC -1- SIMILARITY: TO YEAST EBP1 AND S.POMBE CUT1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M83232; AAA33297.1; ALT_TERM.
DR PIR; A42854; A42854.
DR MEROPS; C50.001; -
KW Nuclear protein; Mitosis.
SQ SEQUENCE 2067 AA; 227925 MW; E0655D939EC148DB CRC64;

Query Match
Best Local Similarity 1.6%; Score 7; DB 1; Length 2067;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 306 GESEHLL 312
|||||||

DB 1417 GESEHLL 1423

RESULT 42
RRPL_NDVYB
ID RRPL_NDVYB STANDARD; PRT; 2204 AA.

AC P11205;
DT 01-JUL-1989 (Rel. 11, Created)


```

Db      157  LPLYNGI 163

RESULT 44
POLG-TEV STANDARD; PRT; 3054 AA.
AC P04517; Q88500; Q88501; Q88504; Q88505; Q88506; Q88773;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DE 15-JUL-1999 (Rel. 38, last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS Tobacco etch virus (TEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Polyviridae;
OC Polyvirus.
OX NCBI_TaxID=12227;
RN [1]
RP SEQUENCE FROM N.A.
RA Allison R., Johnston R.E., Dougherty W.G.;
RT "The nucleotide sequence of the coding region of tobacco etch virus
genomic RNA: evidence for the synthesis of a single polypeptide."
RL Virology 154:9-20(1986).
[2]
RP SEQUENCE OF 2344-3054 FROM N.A.
RA Allison R.E., Sorenson J.C., Kelly M.E., Armstrong F.B.,
RA Dougherty W.G.;
RT "Sequence determination of the capsid protein gene and flanking
RT regions of tobacco etch virus: evidence for synthesis and processing
of a polyprotein in polyvirus genome expression."
RL Proc. Natl. Acad. Sci. U.S.A. 82:3969-3972(1985).
[3]
RP IDENTIFICATION OF PROTEASES.
RX MEDLINE=98251560; Pubmed=2656254;
RA Carrington J.C., Cary S.M., Parks T.D., Dougherty W.G.;
RT "A second proteinase encoded by a plant polyvirus genome."
RL EMBO J. 8:365-370(1989).
CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
MAY BE INVOLVED IN REPLICATION.
CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -1- PPM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
INDIVIDUAL PROTEINS.
CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M15239; AAA47910.1; -
CC EMBL, M1458; AAA47909.1; -
CC DR EMBL; M1216; AAA47908.1; ALT_SEQ.
CC PIR; A04207; GNEVEV.
CC MEROPS; C04.001; -
CC MEROPS; C06.001; -
CC MEROPS; S30.001; -
CC Interpro; IPR001205; -
CC Interpro; IPR001410; -

```

```

DR InterPro; IPR001456; -
DR InterPro; IPR001592; -
DR InterPro; IPR001730; -
DR InterPro; IPR002540; -
DR Pfam; PR002270; DEAD_1.
DR Pfam; PR00863; Peptidase_C4; 1.
DR Pfam; PR00851; Peptidase_C6; 1.
DR Pfam; PR01577; Poly_P1; 1.
DR Pfam; PR00767; Poly-coat; 1.
DR Pfam; PR00680; RNA-dep.RNA-pol; 1.
DR PRINTS; PR00966; NIPOTYPIASE.
KW Hydrolyase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW ATP-binding.
FT CHAIN 1 763 ?
FT CHAIN 2 763 ?
FT CHAIN 3 763 ?
FT CHAIN 4 1163 ?
FT CHAIN 5 1164 1796
FT CHAIN 6 1797 1849
FT CHAIN 7 1850 ?
FT CHAIN 8 2279 ?
FT CHAIN 9 2280 2791
FT CHAIN 10 2792 3054
FT SITE 763 764
FT SITE 1163 1164
FT SITE 1796 1797
FT SITE 1849 1850
FT SITE 2279 2280
FT SITE 2791 2792
FT BINDING 1911 1911
FT NP_BIND 1247 1254
FT SEQUENCE 3054 AA; 346160 MM; 0AF9A3626960B5CE CRC64.
[2]
Query Match 1.6%; Score 7; DB 1; Length 3054;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 23 REARGAR 29
Db 1861 REARGAR 1867
RESULT 45
POLG-LMVO STANDARD; PRT; 3255 AA.
AC P31999; P90263;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS Lettuce mosaic virus (strain 0 / French isolate) (LMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Polyviridae;
OC Polyvirus.
OX NCBI_TaxID=117131;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97239892; Pubmed=9085548;
RA Revers F., Yang S.J., Walter J., Souche S., Lot H., le Gall O.,
RA Candresse T., Dunez J.;
RT "Comparison of the complete nucleotide sequences of two isolates of
RT lettuce mosaic virus differing in their biological properties."
RL Virus Res. 47:167-177(1997).
[2]
RP SEQUENCE OF 2776-3255 FROM N.A.

```

RX MEDLINE-91158469; PubMed-2001176;
 RA Diant S., Lot H., Albouy J., Kuziak C., Meyer M.,
 RA Astier-Manificat S.;
 RT "Nucleotide sequence of the 3' terminal region of lettuce mosaic
 RT potyvirus RNA shows a Gln/Val dipeptide at the cleavage site between
 RT the polymerase and the coat protein.";
 RL Arch. Virol. 116:235-252(1991).
 CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
 CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
 CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
 CC MAY BE INVOLVED IN REPLICATION.
 CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
 CC -1- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
 CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
 CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
 CC INDIVIDUAL PROTEINS.
 CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
 CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
 CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X97704; CAA66280.1; -
 CC EMBL: X65652; CAA66202.1; -
 CC PIR: A43493; A43493.
 CC InterPro: IPR001205; -
 CC InterPro: IPR001410; -
 CC InterPro: IPR001456; -
 CC InterPro: IPR001592; -
 CC InterPro: IPR001730; -
 CC InterPro: IPR002540; -
 CC Pfam: PF00270; DEAD; 1.
 CC Pfam: PF00863; Peptidase_C4; 1.
 CC Pfam: PF00851; Peptidase_C6; 1.
 CC Pfam: PF01577; Poly_P1; 1.
 CC Pfam: PF00767; Poly_coat; 1.
 CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
 CC PRINTS: PR00966; NIAPOTYPYASE.
 CC DR Hydrolyase; Transferase; Thiol protease; RNA-directed RNA polymerase;
 CC DR Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
 CC KW ATP-binding.
 CC KW ATP-binding. 1
 CC FT CHAIN ? ? N-TERMINAL PROTEIN.
 CC FT CHAIN ? ? HELPER COMPONENT PROTEINASE.
 CC FT CHAIN ? ? PROTEIN P3.
 CC FT CHAIN ? ? 6 KDA PROTEIN 1.
 CC FT CHAIN ? ? CYTOPLASMIC INCLUSION PROTEIN.
 CC FT CHAIN ? ? 6 KDA PROTEIN 2.
 CC FT CHAIN ? ? GENOME-LINKED PROTEIN.
 CC FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN A.
 CC FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN B.
 CC FT CHAIN ? ? COAT PROTEIN.
 CC FT CHAIN ? ? ATP (POTENTIAL).
 CC FT NP_BIND 2978 3255
 CC FT NP_BIND 1410 1417
 CC FT SEQUENCE 3255 AA; 367534 MW; CD9AD5E4A0B65953 CRC64;
 SQ

Query Match 1.6%; Score 7; DB 1; Length 3255;
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 LKGCISG 250
 DB 1141 LKGCISG 1147
 RESULT 46
 POLG_LMVE

ID POLG_LMVE STANDARD; PRT; 3255 AA.
 AC P89876;
 DT 15-DEC-1998 (rel. 37, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 15-DEC-1998 (rel. 37, Last annotation update)
 DE GENOME POLYPEPTIDE [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
 DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
 DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
 DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
 DE (EC 3.4.22.-) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
 DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
 DE COAT PROTEIN (CP)].
 OS Lettuce mosaic virus (strain E) (LMV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
 OC Potyvirus.
 OX NCBI_TaxID=117132;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97239892; PubMed-9085548;
 RA Revers F., Yang S.J., Walter J., Souche S., Lot H., Le Gall O.,
 RA Candresse T., Dunez J.;
 RT "Comparison of the complete nucleotide sequences of two isolates of
 RT lettuce mosaic virus differing in their biological properties.";
 RL Virus Res. 47:167-177(1997).
 CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
 CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
 CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
 CC MAY BE INVOLVED IN REPLICATION.
 CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
 CC -1- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
 CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
 CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
 CC INDIVIDUAL PROTEINS.
 CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
 CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
 CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X97705; CAA66281.1; -
 CC MEROPS: C04.001; -
 CC DR InterPro: IPR001205; -
 CC DR InterPro: IPR001410; -
 CC DR InterPro: IPR001456; -
 CC DR InterPro: IPR001592; -
 CC DR InterPro: IPR001730; -
 CC DR InterPro: IPR002540; -
 CC Pfam: PF00270; DEAD; 1.
 CC Pfam: PF00863; Peptidase_C4; 1.
 CC Pfam: PF00851; Peptidase_C6; 1.
 CC Pfam: PF01577; Poly_P1; 1.
 CC Pfam: PF00767; Poly_coat; 1.
 CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
 CC DR Hydrolyase; Transferase; Thiol protease; RNA-directed RNA polymerase;
 CC DR Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
 CC KW ATP-binding.
 CC KW ATP-binding. 1
 CC FT CHAIN ? ? N-TERMINAL PROTEIN.
 CC FT CHAIN ? ? HELPER COMPONENT PROTEINASE.
 CC FT CHAIN ? ? PROTEIN P3.
 CC FT CHAIN ? ? 6 KDA PROTEIN 1.
 CC FT CHAIN ? ? CYTOPLASMIC INCLUSION PROTEIN.
 CC FT CHAIN ? ? 6 KDA PROTEIN 2.
 CC FT CHAIN ? ? GENOME-LINKED PROTEIN.
 CC FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN A.
 CC FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN B.

FT CHAIN ? 2977 NUCLEAR INCLUSION PROTEIN B.
 FT P1AM 2978 3255 COAT PROTEIN.
 FT NP_BIND 1410 1417 ATP (POTENTIAL).
 SO SEQUENCE 3255 AA; 367618 MM; B3E8582927E01628 CRC64;

Query Match 1.68; Score 7; DB 1; Length 3255;
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 244 LKGCISG 250
 DB 1141 LKGCISG 1147

RESULT 47
 POLG_DENS3
 ID POLG_DENS3 STANDARD; PRT; 3390 AA.
 AC P27915;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN NS1, NS2, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)].
 DE DENGE VIRUS TYPE 3.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Flavivirus.
 NC NCB1_Taxid=11069;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90266483; Pubmed=2345967;
 RA Osatomi K., Sumiyoshi H.:
 RT "Complete nucleotide sequence of dengue type 3 virus genome RNA."
 RL Virology 176:643-647(1990).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M9130; AAA99437.1; -
 DR PIR: A34774; GNMVD3.
 DR HSP: P14336; 1SVB.
 DR MEROPS: S07.002; -
 DR InterPro: IPR000069; -
 DR InterPro: IPR000208; -
 DR InterPro: IPR000336; -
 DR InterPro: IPR000404; -
 DR InterPro: IPR000487; -
 DR InterPro: IPR000752; -
 DR InterPro: IPR001122; -
 DR InterPro: IPR001157; -
 DR InterPro: IPR001528; -
 DR InterPro: IPR001850; -
 DR InterPro: IPR002535; -
 DR Pfam: PF01004; Flav1_NS1; 1.
 DR Pfam: PF00948; Flav1_NS1; 1.
 DR Pfam: PF01005; Flav1_NS2a; 1.
 DR Pfam: PF01002; Flav1_NS2b; 1.
 DR Pfam: PF01350; Flav1_NS4A; 1.
 DR Pfam: PF01349; Flav1_NS4B; 1.
 DR Pfam: PF00972; Flav1_NS5; 1.
 DR Pfam: PF01003; Flav1_capsid; 1.
 DR Pfam: PF00869; Flav1_glycoprot; 1.

DR Pfam: PF00949; Flav1_helicase; 1.
 DR Pfam: PF01570; Flav1_propep; 1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein.
 FT CHAIN 1 114
 FT PROPEP 115 205
 FT CHAIN 206 280
 FT CHAIN 281 773
 FT CHAIN 774 1184
 FT CHAIN 1185 1343
 FT CHAIN 1344 1473
 FT CHAIN 1474 2092
 FT CHAIN 2093 2378
 FT CHAIN 2379 2490
 FT CHAIN 2491 3390
 FT NP_BIND 1667 1674
 FT SITE 1758 1761
 FT TRANSMEM 46 67
 FT TRANSMEM 266 280
 FT TRANSMEM 724 746
 FT TRANSMEM 753 771
 FT TRANSMEM 1156 1175
 FT DISULFID 283 310
 FT DISULFID 340 396
 FT DISULFID 354 385
 FT DISULFID 372 401
 FT DISULFID 463 563
 FT DISULFID 580 611
 FT CARBOHYD 183 183
 FT CARBOHYD 347 347
 FT CARBOHYD 433 433
 FT CARBOHYD 750 750
 FT CARBOHYD 903 903
 FT CARBOHYD 980 980
 FT CARBOHYD 1132 1132
 FT CARBOHYD 1188 1188
 FT CARBOHYD 1661 1661
 FT CARBOHYD 2300 2300
 FT CARBOHYD 2304 2304
 FT CARBOHYD 2456 2456
 FT CARBOHYD 2702 2702
 FT CARBOHYD 2712 2712
 SO SEQUENCE 3390 AA; 378057 MM; 666EB870E1E1756E CRC64;
 Query Match 1.68; Score 7; DB 1; Length 3390;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 OY 43 AAVLLV 49
 DB 2351 AAVLLV 2357
 RESULT 48
 VC38_BPM2
 ID VC38_BPM2 STANDARD; PRT; 50 AA.
 AC O64229;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GENE 38 PROTEIN (GP38).
 GN 38.
 OS Mycobacteriophage D29.
 OC Viruses.
 OX NCB1_Taxid=28369;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96300335; Pubmed=9636706;
 RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
 RT "Genome structure of mycobacteriophage D29: implications for phage evolution."

```

RL J. Mol. Biol. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF022214; AAC18479.1; -
DR SEQUENCE 50 AA; 4851 MW; 75BC1A1CF2EF26E CRC64;
SO

```

Query Match 1.4%; Score 6; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 44 AVLLV 49
    |||||
DB 5 AVLLV 10

```

RESULT 49

```

HSP1_MACRU
ID HSP1_MACRU STANDARD; PRT; 59 AA.
AC P42142;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPERM PROTAMINE PL.
GN PRML.
OS Macropus rufus (Red kangaroo) (Megaloptera rufa).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9321;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sperm;
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfeld R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond. B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L35447; AAA74616.1; -
DR InterPro: IPR000221;
DR Pfam: PF00260; protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MER 0
FT BY SIMILARITY.
SO SEQUENCE 59 AA; 8230 MW; 78F1AE592B4B2F2A2 CRC64;

```

Query Match 1.4%; Score 6; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 RSQRRR 324
|||||

```

DB 25 RSQRRR 30

```

RESULT 50

```

HSP1_PHACI
ID HSP1_PHACI STANDARD; PRT; 59 AA.
AC P79990;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SPERM PROTAMINE PL.
GN PRML.
OS Phascolarctos cinereus (Koala).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phascolarctidae; Phascolarctos.
OX NCBI_TaxID=38626;
RN [1]
RP SEQUENCE FROM N.A.
RA Retief J.D., Krajewski C., Westernman M., Winkfeld R.J., Dixon G.H.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U87789; AAB88911.1; -
DR InterPro: IPR000221;
DR Pfam: PF00260; protamine_P1; 1.
DR PROSITE: PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MER 0
FT BY SIMILARITY.
SO SEQUENCE 59 AA; 8271 MW; 4E1FB24E21952EA2 CRC64;

```

Query Match 1.4%; Score 6; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 RSQRRR 324
|||||

DB 25 RSQRRR 30

Search completed: May 23, 2001, 14:22:32
Job time: 116 sec

This Page Blank (uspto)